

Empirical Mode Decomposition:
A useful technique for neuroscience?

**Review of Huang et al, Proc. R. Soc. Lond. A
(1998) v. 454, 903-995**

**By Robert Liu
Computational Journal Club
January 11, 2002**

Main question

How can we analyze non-stationary and nonlinear time series?

- Stationary time series: ensemble means are time-shift invariant
 - Weakly stationary: true for 1st and 2nd order means

$$\left. \begin{aligned} E(|X(t)|^2) &< \infty, \\ E(X(t)) &= m, \\ C(X(t_1), X(t_2)) &= C(X(t_1 + \tau), X(t_2 + \tau)) = C(t_1 - t_2), \end{aligned} \right\} \quad (1.1)$$

- A non-stationary example is a transient signal like a delta function
- Nonlinear time series: generated by an underlying dynamical system obeying nonlinear equations
 - Often linear approximations can be made -> complex time series can be decomposed into a superposition of simple solutions (e.g. sinusoidal waves)

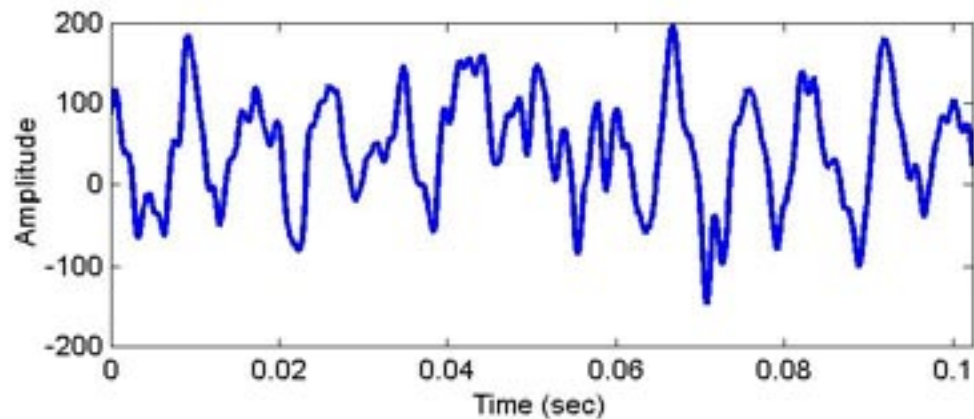
What do we want to gain from the analysis?

- Determine characteristic time / frequency scales for the energy

Motivation

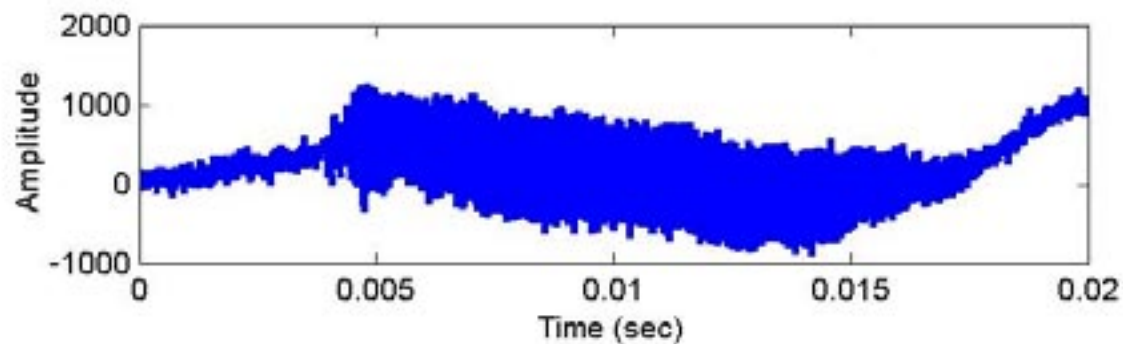
Neuroscience is filled with complex time series

- Local field potentials and electroencephalograms



LFP from
Cat K28

- Complex natural stimuli



Mouse pup call

Existing methods

Dominant approach: decomposition of the time series into component basis functions satisfying

- (a) Completeness of the basis
- (b) Orthogonality of the basis

Examples:

- Fourier methods
- Wavelet analysis
- Principal component analysis
- Etc. . . .

Fourier methods

Method

- Decompose time series, $f(t)$, into global sinusoidal components of fixed amplitude, a_j , (e.g. FFT) -> complete & orthogonal basis

$$f(t) = \sum_{j=0}^n a_j e^{i\omega_j t}, \quad a_j = \frac{1}{2\pi} \int_t f(t) e^{-i\omega_j t} dt. \quad (8.3)$$

Interpretation

- The spectral amplitudes, a_j , yield the energy contributed by a sinusoid at frequency ω_j that spans the whole time series

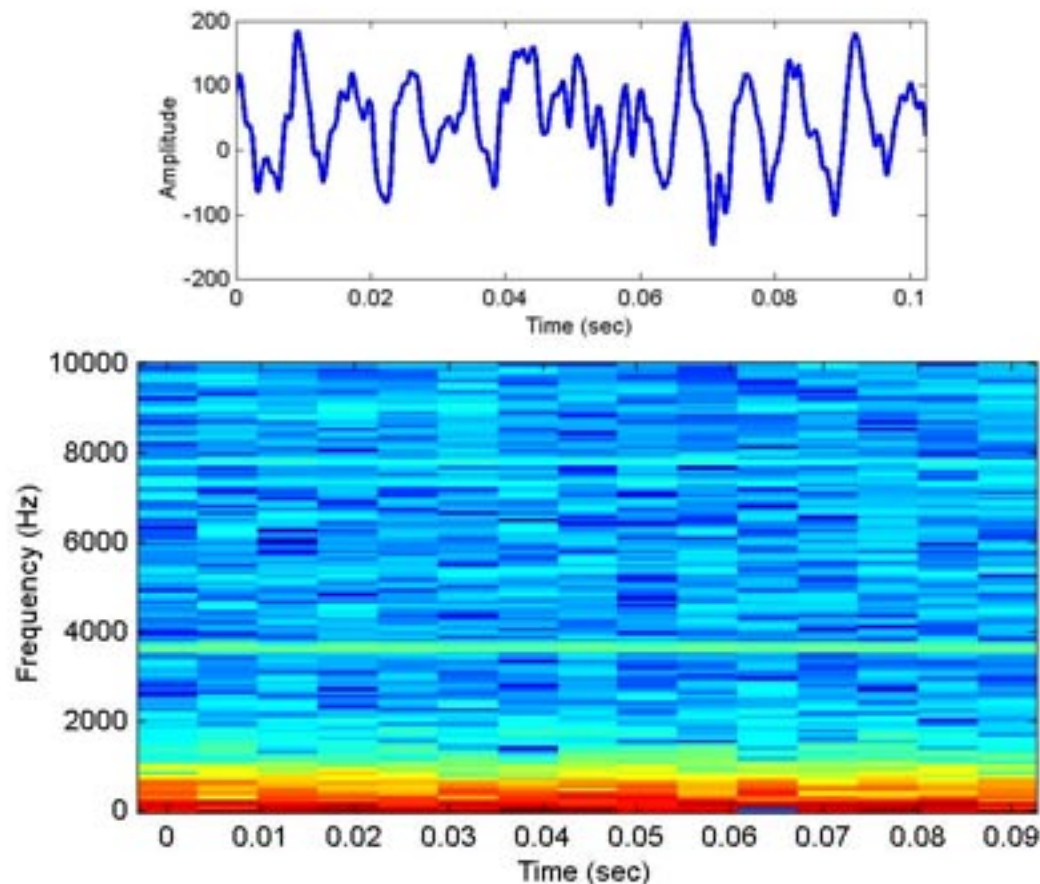
Most useful when

- The underlying process is linear so that the superposition of sinusoidal solutions makes physical sense
- The time series is stationary, since the a_j are constant
 - If not, the spectral energy spreads, often requiring carefully-phased, global (possibly harmonic) sinusoidal components to reconstruct the non-uniform time series
 - The Fourier spectrum loses track of time location for events -> not a local description

Fourier methods

If piece-wise stationarity can be assumed

- Construct spectrogram by sliding a window across the time-series and performing Fourier analysis



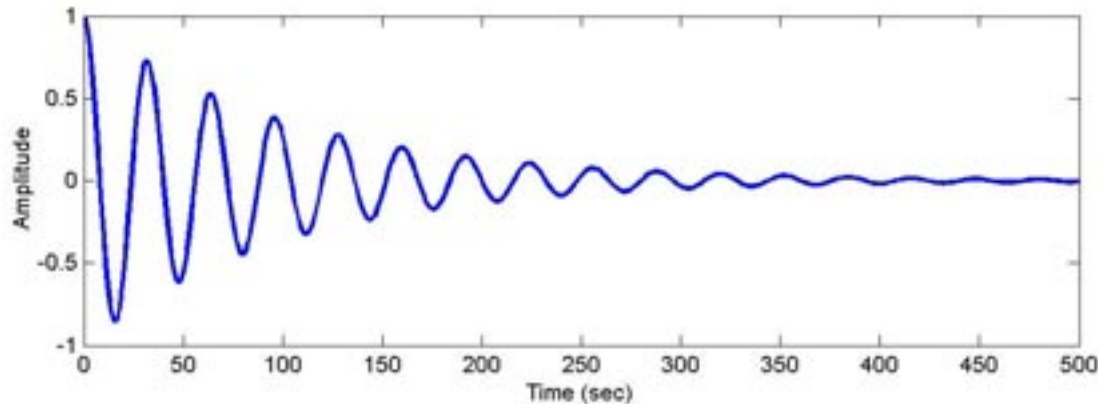
- Constrained by a time-frequency tradeoff -> often don't capture the temporal events at the desired frequency resolution -> still nonlocal

Fourier methods

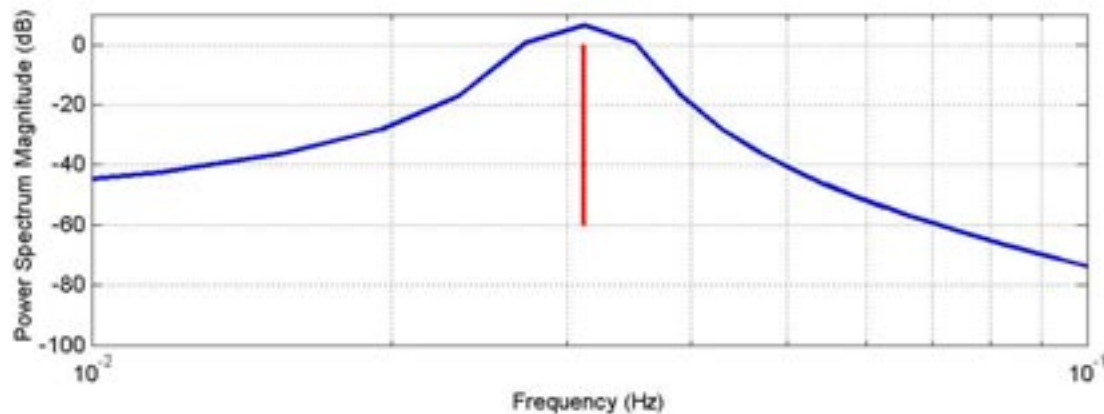
But even if weak stationarity holds, Fourier methods may not give an efficient (adaptive) representation

- Single decaying sinusoid is compactly described as:

$$x = \exp(-0.01t) \cos \frac{2}{32}\pi t, \quad (8.9)$$



Note this is not stationary if it starts abruptly at $t=0$



But Fourier representation is spread over frequency

Wavelet analysis

Method

- Decompose time series, $X(t)$, into local, time-dilated and time-translated wavelet components, $\psi \rightarrow$ complete (not necessarily orthogonal) basis

$$W(a, b; X, \psi) = |a|^{-1/2} \int_{-\infty}^{\infty} X(t) \psi^* \left(\frac{t - b}{a} \right) dt, \quad (2.1)$$

Interpretation

- W represents the energy in X of temporal scale a at $t=b$

Attractive because. . . .

- Local, although higher frequencies are more localized
- Uniform temporal resolution for all frequency scales
 - But, resolution is limited by the basic wavelet
- Useful for characterizing gradual frequency changes

But. . . .

- Non-adaptive -- same basic wavelet is used for all data

Principal component analysis

Method

- Decompose time series, $z(t)$, into eigenbases of the covariance matrix -> complete, orthogonal basis

$$z(x, t) = \sum_1^n a_k(t) f_k(x), \quad f_j \cdot f_k = \delta_{jk}, \quad C \cdot f_k = \lambda_k f_k,$$

Interpretation

- Modes f_j often interpreted as “directions” of independent variations, but they may not be physical modes

Attractive because. . . .

- Adaptive -- derived from the data

But. . . .

- Distribution of eigenvalues do not yield characteristic time or frequency scales
- Eigenmodes themselves need not be linear or stationary (and therefore are still not easily analyzed by spectral methods)

What would we like?

A decomposition that is

- (a) Complete
- (b) Orthogonal
- (c) Local
- (d) Adaptive

From which local time / frequency scales are extracted

How do we define local time / frequency scales?

- Instantaneous frequency

Instantaneous vs. global frequency

Global frequency = average frequency derived by weighting by the Fourier power spectrum, $|S(\omega)|^2$

$$\langle \omega \rangle = \int \omega |S(\omega)|^2 d\omega,$$

Instantaneous frequency of a signal, $X(t)$

- Form an analytic signal, $Z(t)$, from $X(t)$
 - Take the Hilbert transform of $X(t)$

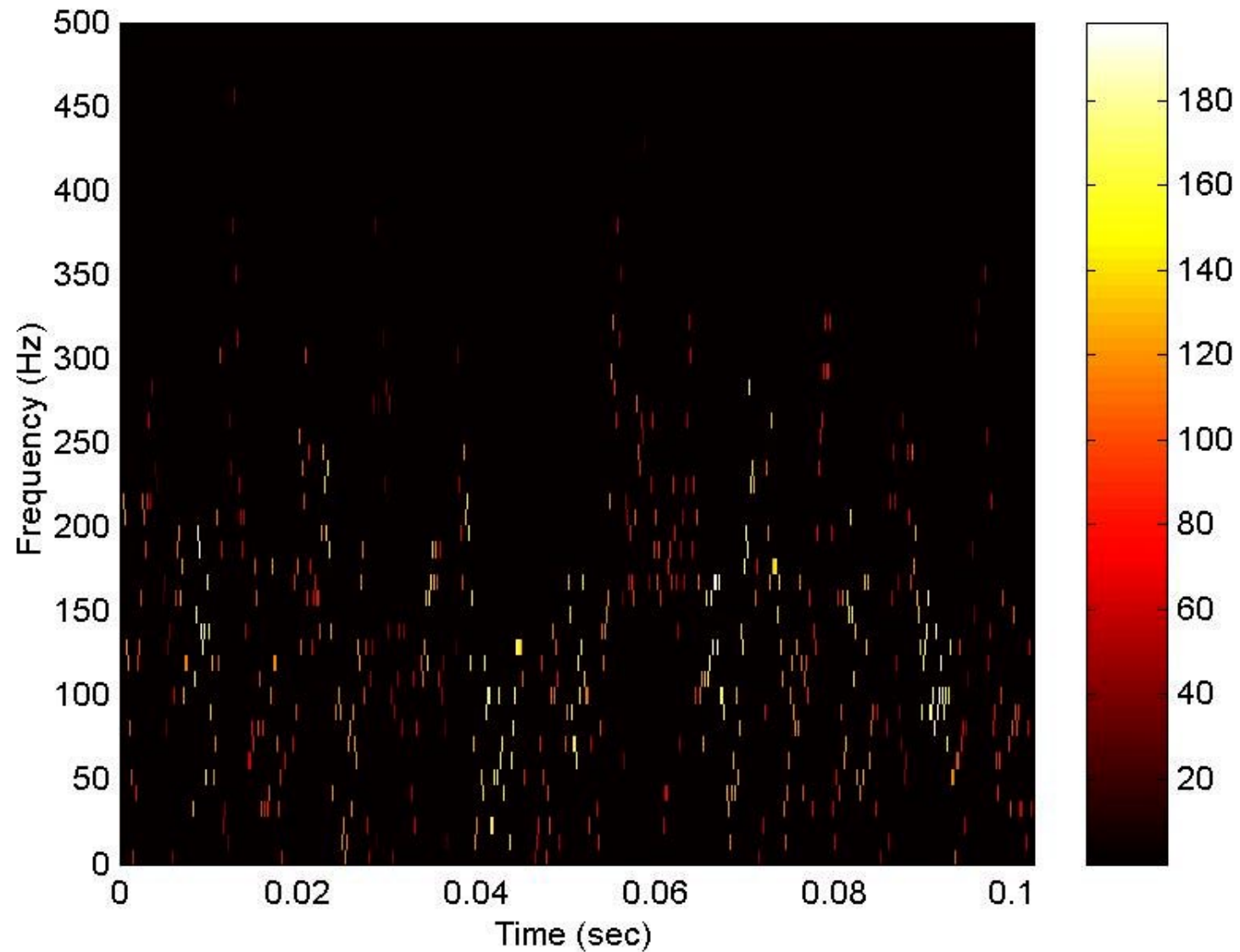
$$Y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{X(t')}{t - t'} dt', \quad Z(t) = X(t) + iY(t) = a(t)e^{i\theta(t)}$$

- Physical interpretation: $Y(t)$ is the best local fit (since the $1/t$ emphasizes local properties) of a trig function to $X(t)$
 - Polar coordinate description for $Z(t)$ allows the phase $\theta(t)$ to be defined
- $Z(t)$ has the same positive frequency spectrum as $X(t)$, but zero negative frequency spectrum
- Define instantaneous frequency as the rate of phase change

$$\omega = \frac{d\theta(t)}{dt}.$$

Hilbert spectrum of the LFP

Not so useful. . . .



Valid description?

When is such a description valid?

- If this concept is applied blindly to any analytic function, one may run into one of 4 paradoxes (L. Cohen)
 - 1) “The instantaneous frequency may not be one of the frequencies in the (Fourier) spectrum.”
 - 2) “If we have a line spectrum consisting of only a few sharp frequencies, then the instantaneous frequency may be continuous and range over an infinite number of values.”
 - 3) “Although the spectrum of the analytic signal is zero for negative frequencies, the instantaneous frequency may be negative.”
 - 4) “For a bandlimited signal the instantaneous frequency may go outside the band.”
- Instead, the definition implies that at a given time, there is only a SINGLE frequency -> “mono-component”
 - Perhaps by restricting to “narrow-bandwidth” signals, but attempts to define this are vague, and “global”
 - Global restrictions have been specified which allow for a meaningful instantaneous frequency: “real part of Fourier transform has to have only positive frequency”

A simple example

What signal
properties cause
difficulties in the
instantaneous
frequency?
Consider. . . .

$$X(t) = \alpha + \sin t$$

$$Y(t) = \cos t$$

$$Z(t) = a + \sin t + i \cos t$$

For. . . .

$\alpha = 0$, const. freq, OK

Bad signals | $\alpha < 1$, dc., asym. env.
| $\alpha > 1$, neg. freq, riding
| wave

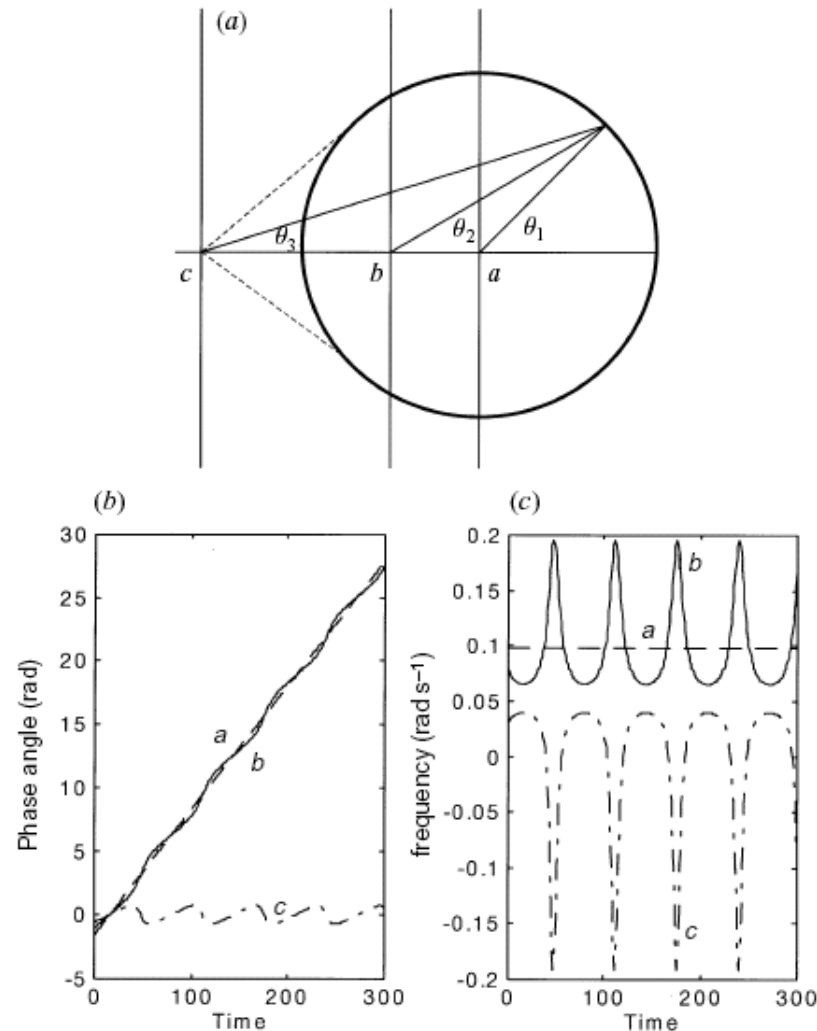


Figure 1. Physical interpretation of instantaneous frequency. (a) The phase plane for the model functions of $x(t) = \alpha + \sin t$. (a) $\alpha = 0$; (b) $\alpha < 1$; (c) $\alpha > 1$. (b) The unwrapped phase function of the model functions. (c) The instantaneous frequency computed according to equation (3.4).

Goals of the new method

1) Decompose time series into superposition of components with well defined instantaneous frequency -> Intrinsic Mode Functions (IMF)

- Components should (approximately) obey earlier requirements of completeness, orthogonality, locality and adaptiveness.
- To get an IMF, need to
 - LOCALLY eliminate riding waves
 - LOCALLY eliminate asymmetries (defined by envelope of extrema)

2) Construct the Hilbert spectrum of each IMF, representing it in the amplitude - instantaneous frequency - time plane

Properties of IMF's

An intrinsic mode function (IMF) is a function that satisfies two conditions: (1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; and (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

- (1) corresponds loosely to finding “narrow-band” signals, or eliminating “riding-waves”**
- (2) ensures that the instantaneous frequency will not have fluctuations arising from an asymmetric wave forms**

The name ‘intrinsic mode function’ is adopted because it represents the oscillation mode imbedded in the data. With this definition, the IMF in each cycle, defined by the zero crossings, involves only one mode of oscillation, no complex riding waves are allowed. With this definition, an IMF is not restricted to a narrow band signal, and it can be both amplitude and frequency modulated. In fact, it can be non-stationary. As

Both imply finding modes with zero LOCAL mean

that the components all satisfy the conditions imposed on them. Physically, the necessary conditions for us to define a meaningful instantaneous frequency are that the functions are symmetric with respect to the *local zero mean*, and have the same numbers of zero crossings and extrema. Based on these observations, we propose a class

Finding IMFs: Empirical Mode Decomposition

Most data are not naturally IMF's

starting point. Unfortunately, most of the data are not IMFs. At any given time, the data may involve more than one oscillatory mode; that is why the simple Hilbert transform cannot provide the full description of the frequency content for the general data as reported by Long *et al.* (1995). We have to decompose the data into IMF

What are we looking for?

The essence of the method is to identify the intrinsic oscillatory modes by their characteristic time scales in the data empirically, and then decompose the data

Assumptions

- 1) At least two extrema -- one max, one min
- 2) Characteristic time scale defined by the time between extrema
- 3) If no extrema, differentiation will reveal extrema (integrate at end)

The sifting process

Procedure

- Identify the extrema (both maxima and minima) of the data, $X(t)$
- Generate the envelope by connecting maxima points with a cubic spline, and minima points with a cubic spline
- Determine the LOCAL mean, m_1 , by averaging the envelope
- Since IMF should have zero local mean, subtract out the mean from the data

$$X(t) - m_1 = h_1. \quad (5.1)$$

- h_1 is probably not an IMF; repeat as necessary until it is
- -> End up with an IMF

As described above, the process is indeed like sifting: to separate the finest local mode from the data first based only on the characteristic time scale. The sifting process, however, has two effects: (a) to eliminate riding waves; and (b) to smooth uneven amplitudes.

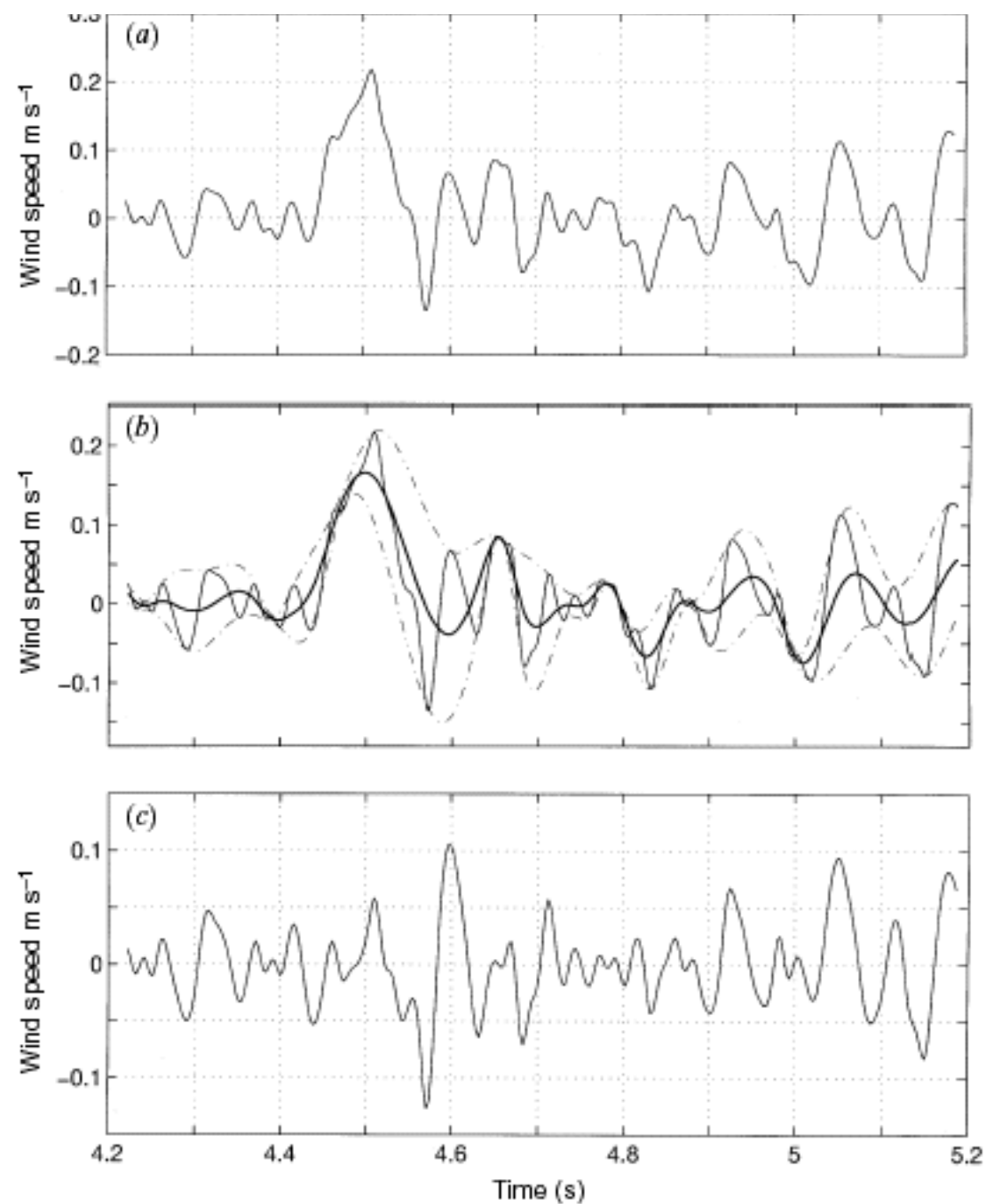


Figure 3. Illustration of the sifting processes: (a) the original data; (b) the data in thin solid line, with the upper and lower envelopes in dot-dashed lines and the mean in thick solid line; (c) the difference between the data and the mean of the envelopes. This is still not an IMF, for these oscillations level

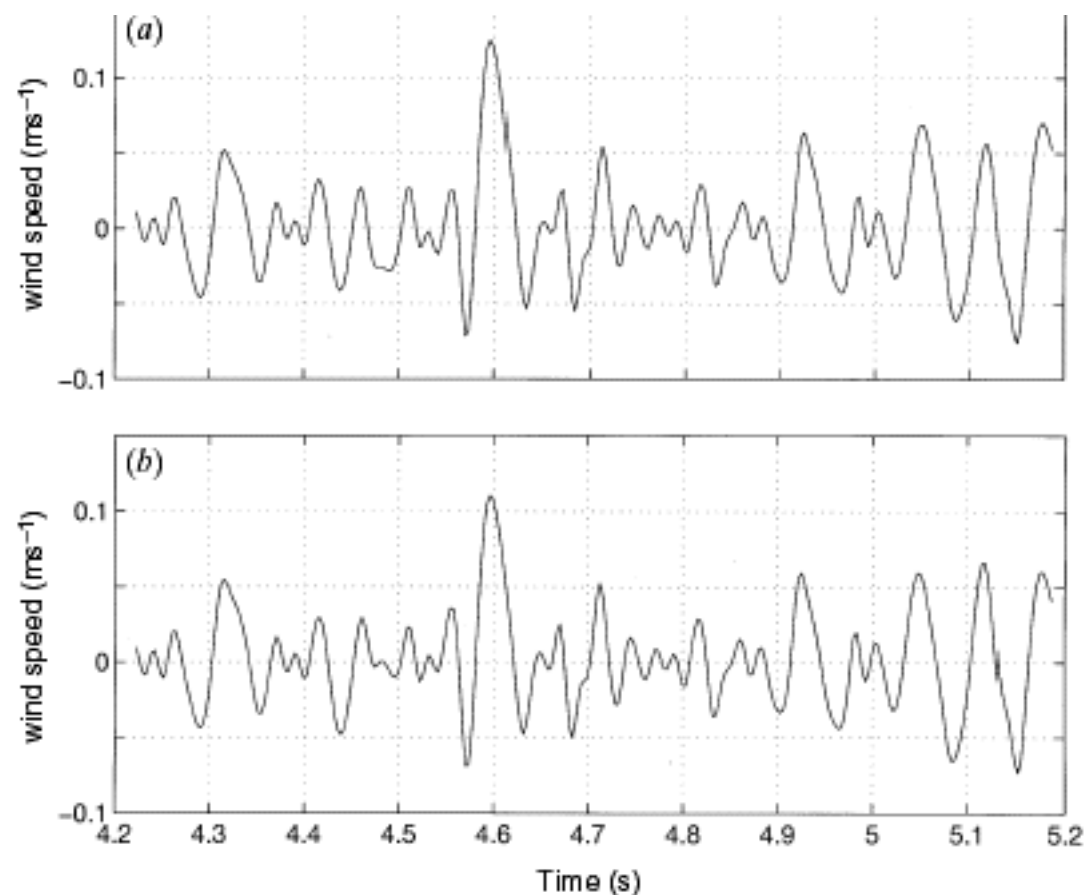
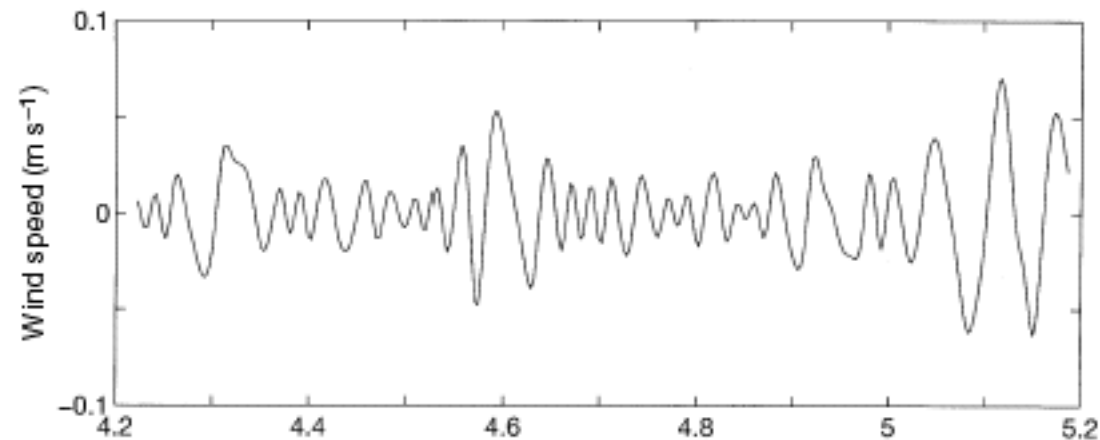


Figure 4. Illustration of the effects of repeated sifting process: (a) after one more sifting of the result in figure 3c, the result is still asymmetric and still not a IMF; (b) after three siftings, the result is much improved, but more sifting needed to eliminate the asymmetry. The final IMF is show



Finding all the IMFs

Procedure

- Once the first set of “siftings” results in an IMF, define

$$c_1 = h_{1k}, \quad (5.4)$$

- This 1st component contains the finest temporal scale in the signal
- Generate the residue, r_1 , of the data by subtracting out c_1

$$X(t) - c_1 = r_1. \quad (5.6)$$

- The residue now contains information about longer periods -> resift to find additional components

$$r_1 - c_2 = r_2, \dots, r_{n-1} - c_n = r_n. \quad (5.7)$$

- Form the superposition of the components to reconstruct the data

$$X(t) = \sum_{i=1}^n c_i + r_n. \quad (5.8)$$

- Linear superposition of modes
- Efficient/adaptive (based on data) representation

Some details. . . .

When does the sifting stop?

Therefore, the sifting process should be applied with care, for carrying the process to an extreme could make the resulting IMF a pure frequency modulated signal of constant amplitude. To guarantee that the IMF components retain enough physical sense of both amplitude and frequency modulations, we have to determine a criterion for the sifting process to stop. This can be accomplished by limiting the size of the standard deviation, SD, computed from the two consecutive sifting results as

$$\text{SD} = \sum_{t=0}^T \left[\frac{|(h_{1(k-1)}(t) - h_{1k}(t))|^2}{h_{1(k-1)}^2(t)} \right]. \quad (5.5)$$

A typical value for SD can be set between 0.2 and 0.3. As a comparison, the two Fourier spectra, computed by shifting only five out of 1024 points from the same data, can have an equivalent SD of 0.2–0.3 calculated point-by-point. Therefore, a SD value of 0.2–0.3 for the sifting procedure is a very rigorous limitation for the difference between siftings.

- In practice, this criterion seems to “work”
- Drawback: ad hoc criterion

Technical problems

- Cubic spline fitting is computationally intensive and creates distortions near the end points

Example of a turbulence data set

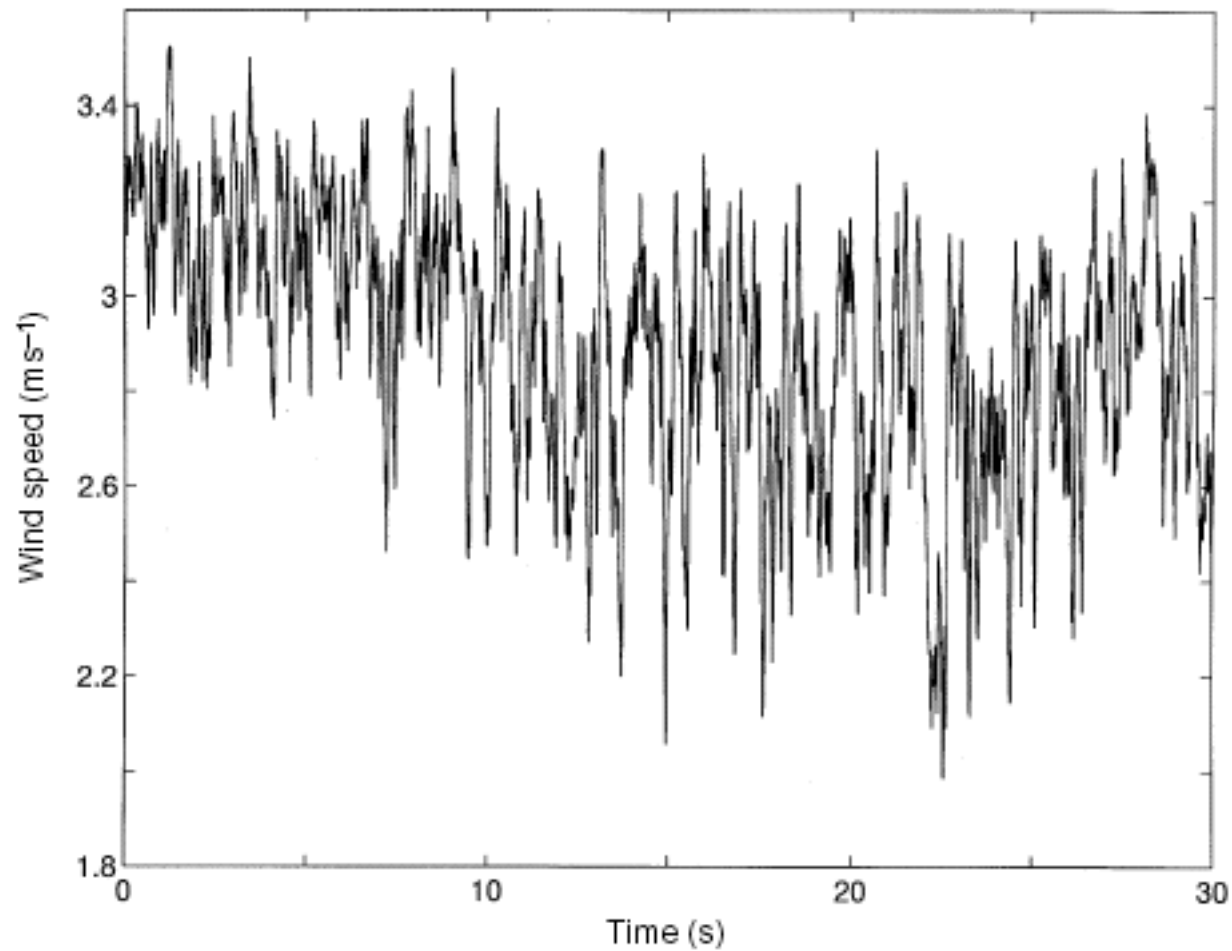
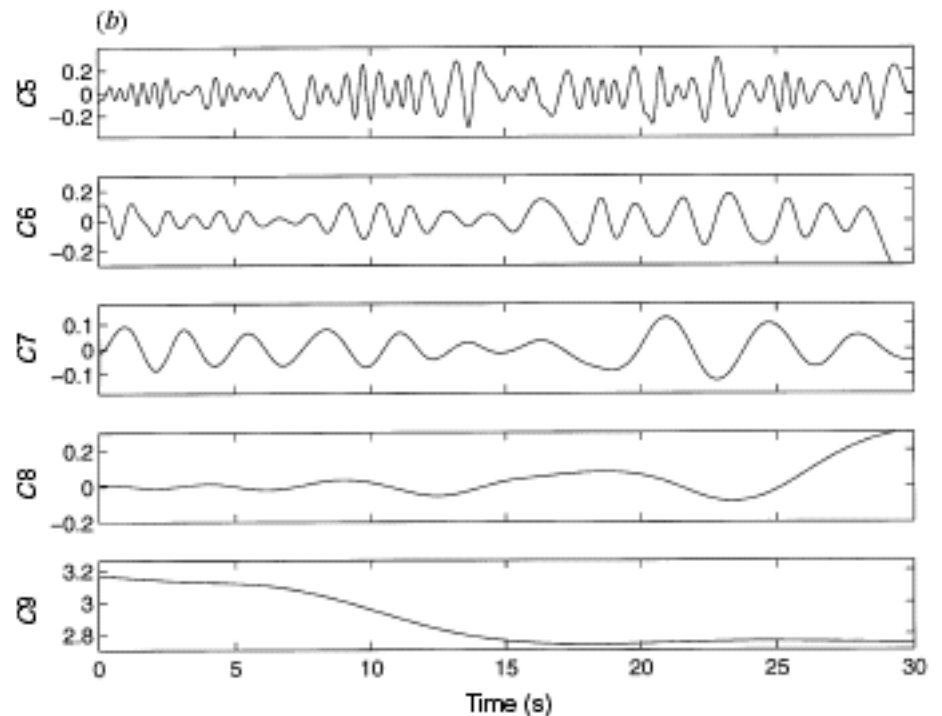
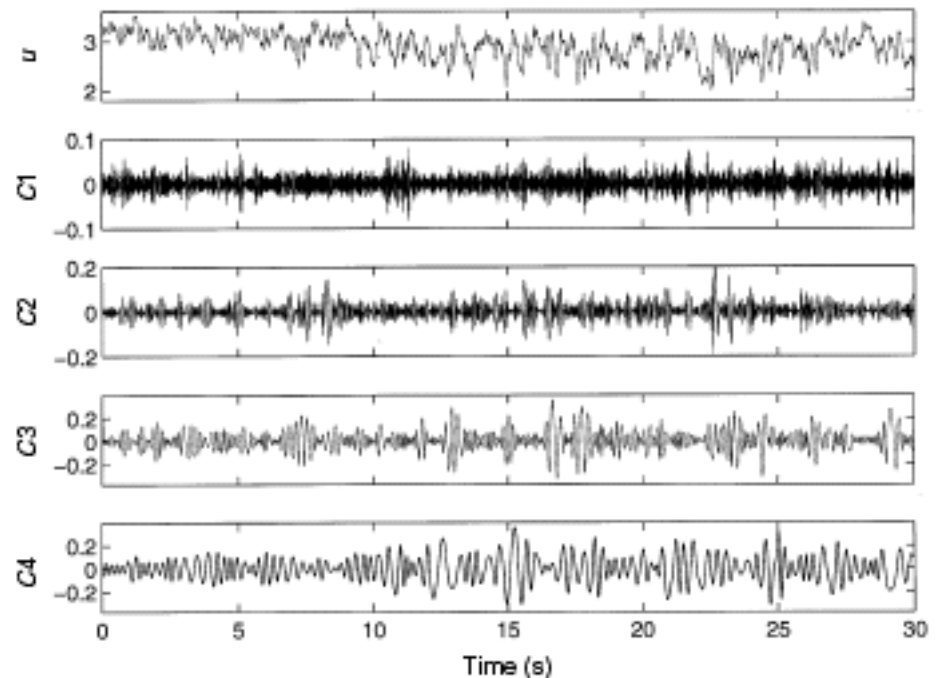


Figure 5. Calibrated wind data from a wind-wave tunnel.

Decomposition

Total of 9 components

- Efficient representation (relative to Fourier) of a turbulence data set
- Never have oscillations of the same scale at the same time in different modes



Satisfying our criteria

4) Adaptive

- By construction, since based on the data

3) Local

- By construction, since local properties of the oscillatory modes are emphasized

1) Completeness

- Complete for a given data set, as established by Eqn. (5.8)
- Check numerically by reconstructing the data (see next page)

2) Orthogonality

- Not guaranteed theoretically, although by construction, the components should be locally orthogonal
- Orthogonality prevents energy leakage between modes
- Check a posteriori with the index of orthogonality (IO)

$$\text{IO} = \sum_{t=0}^T \left(\sum_{j=1}^{n+1} \sum_{k=1}^{n+1} C_j(t) C_k(t) / X^2(t) \right). \quad (6.4)$$

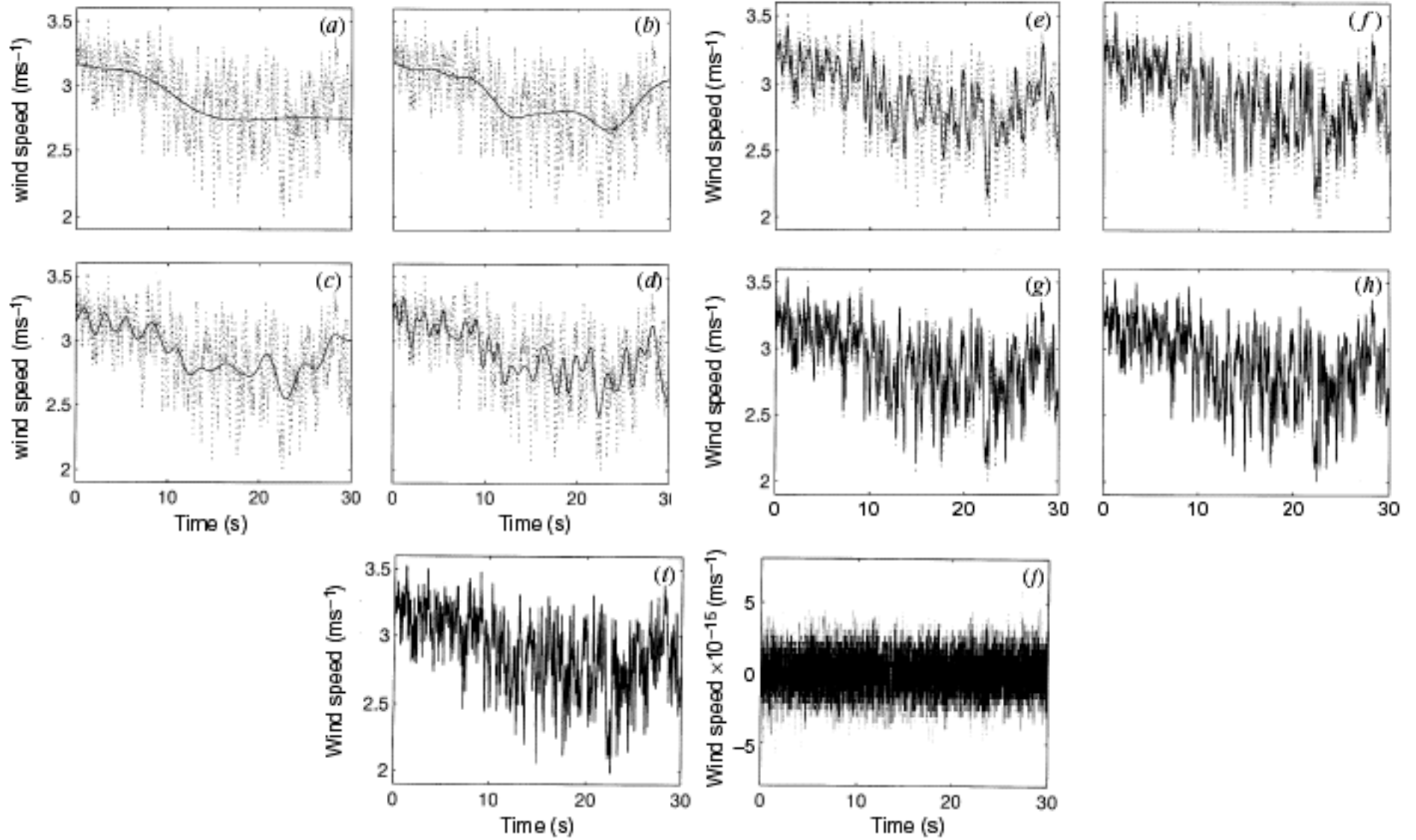


Figure 7. *Cont.* (e) Data (in the dotted line) and the sum of c_9 – c_5 components (in the solid line). (f) Data (in the dotted line) and the sum of c_9 – c_4 components (in the solid line). (g) Data (in the dotted line) and the sum of c_9 – c_3 components (in the solid line). By now, we seem to have recovered all the energy containing eddies. (h) Data (in the dotted line) and the sum of c_9 – c_2 components (in the solid line). (i) Data (in the dotted line) and the sum of c_9 – c_1 components (in the solid line). This is the final reconstruction of the data from the IMFs. It appears no different from the original data. (j) The difference between the original data and the reconstructed one;

Visualizing the data

Hilbert spectrum

- In terms of the IMFs, the time series can now be written as

$$X(t) = \sum_{j=1}^n a_j(t) \exp \left(i \int \omega_j(t) dt \right). \quad (7.1)$$

- This is a generalization of the Fourier decomposition, allowing for time varying amplitudes and frequencies, thus simplifying the description of non-stationary data
- The spectrum describes the joint distribution of the amplitude and frequency content of the signal as a function of time
 - Often presented as contour or color plots of the amplitude (or energy) over the time-frequency plane
 - Time resolution can be as precise as the sampling period
 - Frequency resolution is arbitrary (up to the Nyquist frequency)
 - Lowest extractable frequency (best resolution) = $1 / T$ Hz, T = duration
 - Highest extractable frequency = $1 / (n \, dt)$ Hz, dt = sample period, n = minimum number of pts to accurately define frequency = 5
 - -> Max number of bins = $T / (n \, dt)$

Hilbert spectrum

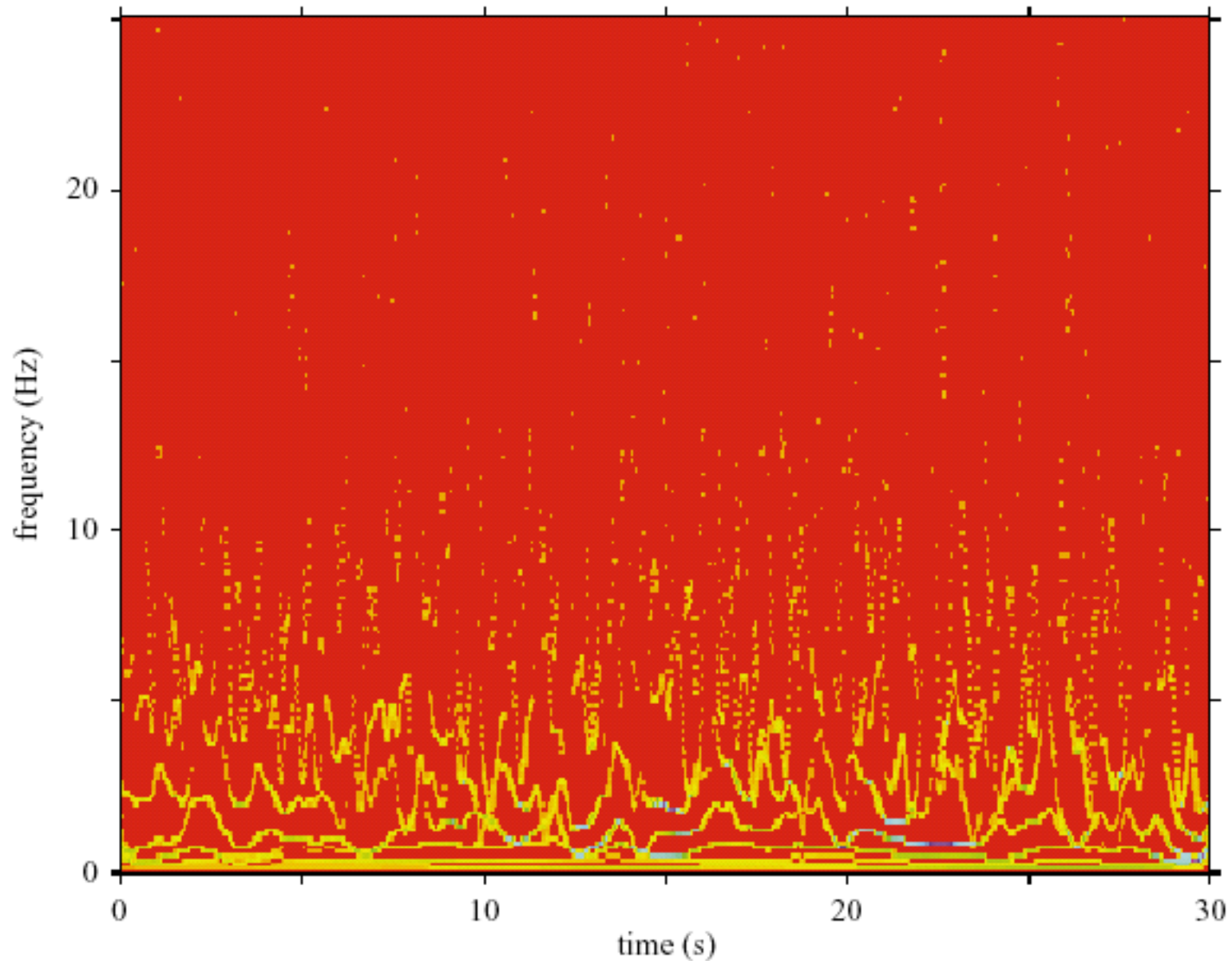


Figure 8. The Hilbert spectrum for the wind data with 200 frequency cells. The wind energy appears in skeleton lines representing each IMF

Morlet wavelet spectrum

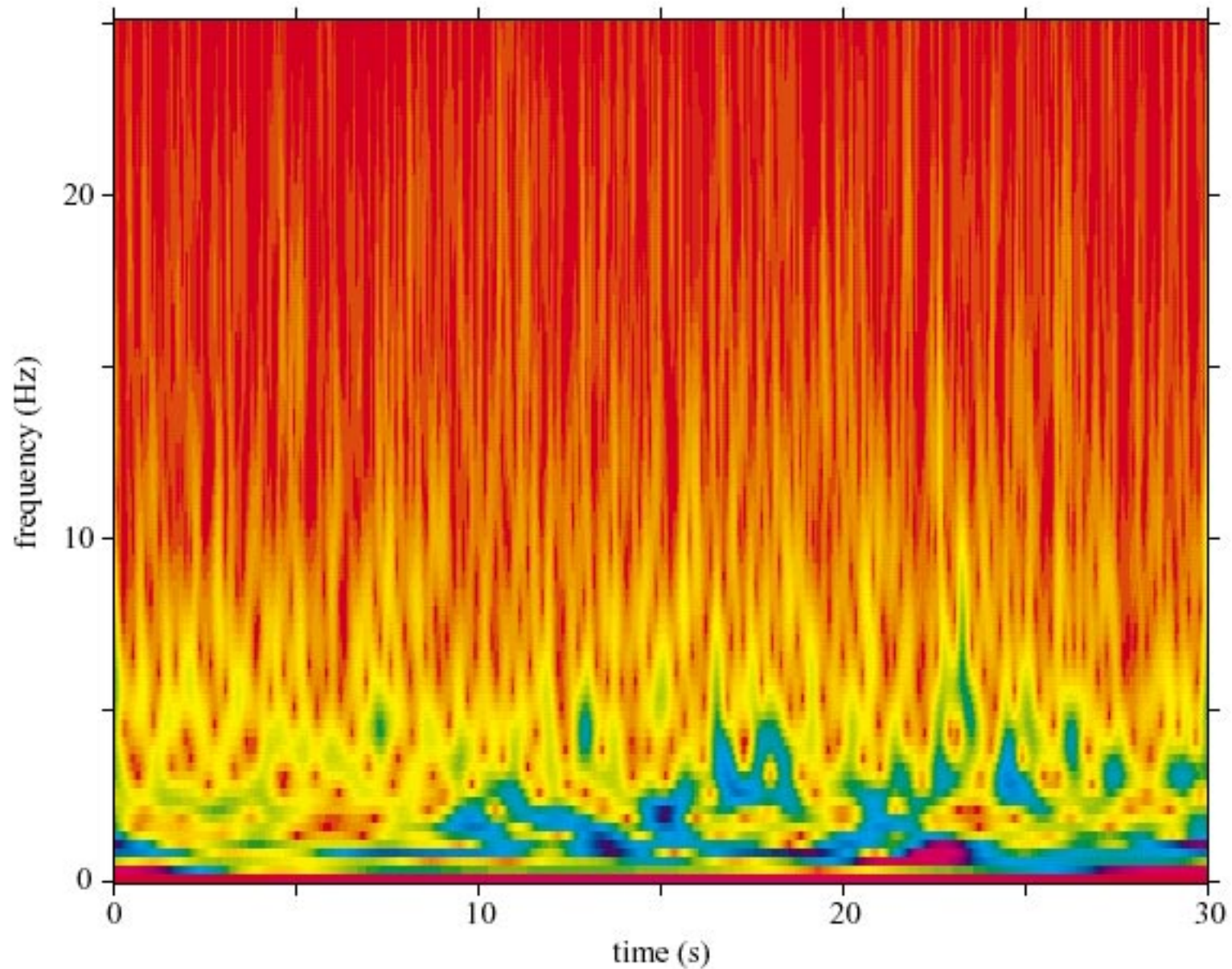


Figure 9. The Morlet wavelet spectrum for the wind data with the same number of frequency cells. Wind energy appears in smoothed contours with a rich energy distribution in the high

Smoothed Hilbert spectrum

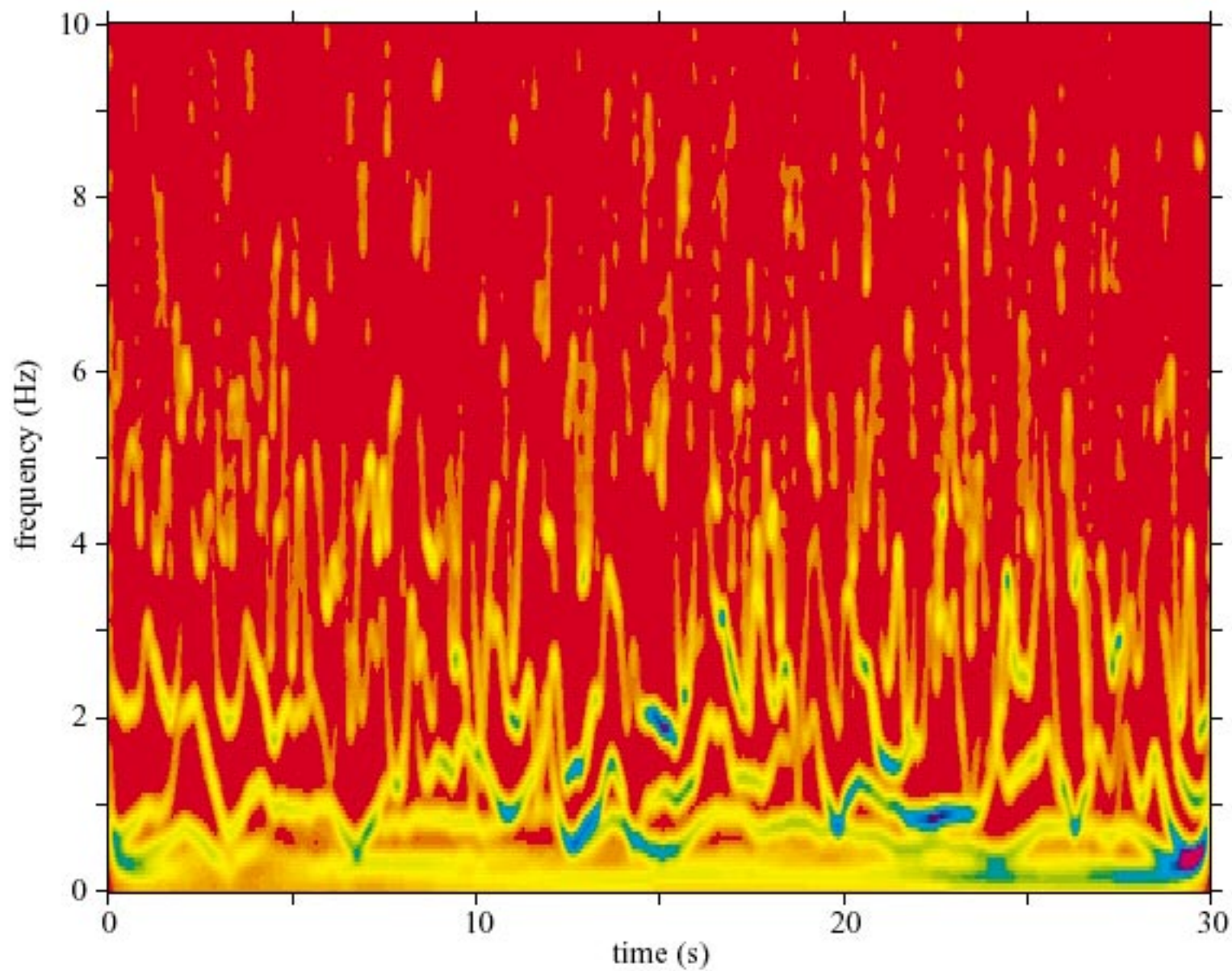


Figure 10. A 15×15 weighted Gaussian filtered Hilbert spectrum. The appearance gives a better comparison with the wavelet result, but the information content has been degraded.

Quantifying stationarity

Simple idea: Stationary processes should have horizontal contours in the Hilbert spectrum

- Define the marginal Hilbert spectrum by integrating over time:

$$h(\omega) = \int_0^T H(\omega, t) dt. \quad (7.4)$$

- Define the mean marginal spectrum as

$$n(\omega) = \frac{1}{T} h(\omega).$$

- Degree of stationarity over the whole data set is then

$$DS(\omega) = \frac{1}{T} \int_0^T \left(1 - \frac{H(\omega, t)}{n(\omega)} \right)^2 dt, \quad (7.7)$$

- The closer to zero this is, the more stationary the system

Calibrating the technique: single cycle

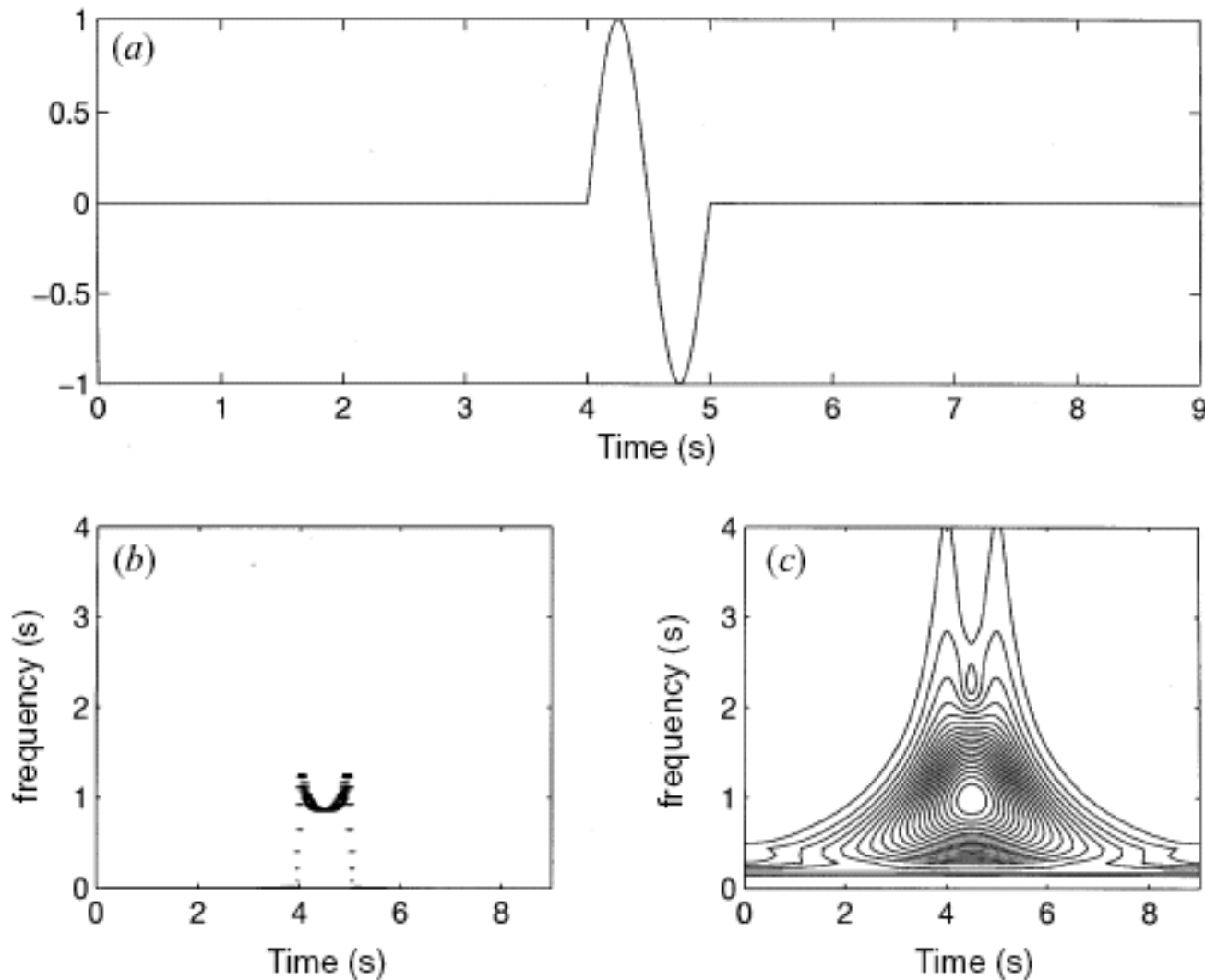


Figure 14. A calibration of time localization of the Hilbert spectrum analysis. (a) The calibration data, a single sine wave. (b) The Hilbert spectrum for the calibration signal: the energy is highly localized in time and frequency, though there are some end effects. (c) The Morlet wavelet spectrum for the calibration signal: the calibration signal is localized by the high-frequency components, yet the energy distribution in the frequency space spreads widely in comparison

Calibrating the technique: Stokes wave

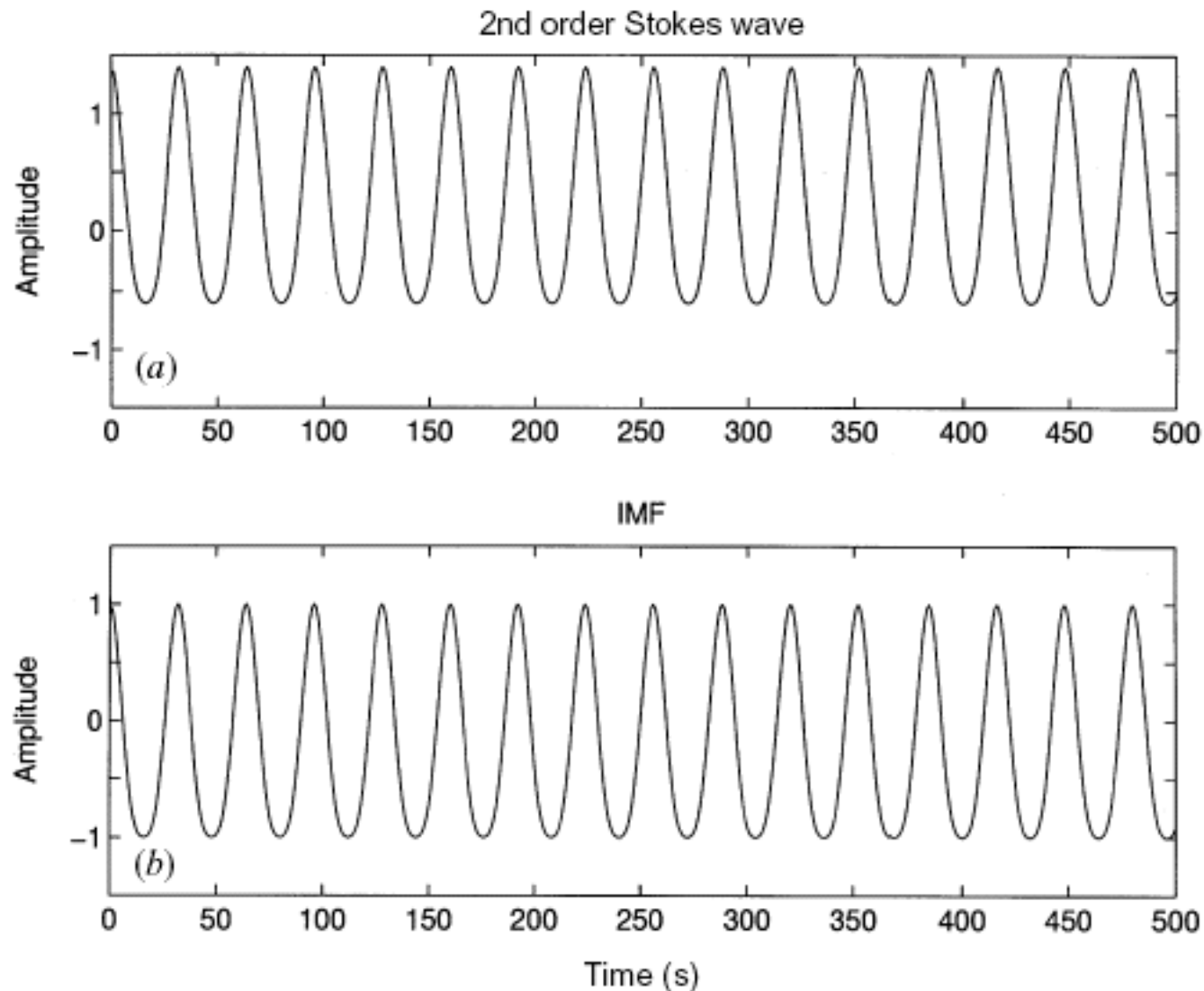


Figure 15. Validation of the intrawave frequency modulation with Stokian waves. (a) The profile of a second-order Stokes wave in deep water with sharp crests and rounded-off troughs in comparison with the pure cosine waves. (b) The IMF generated by the Stokes wave, there is only one component; the constant off-set is not shown.

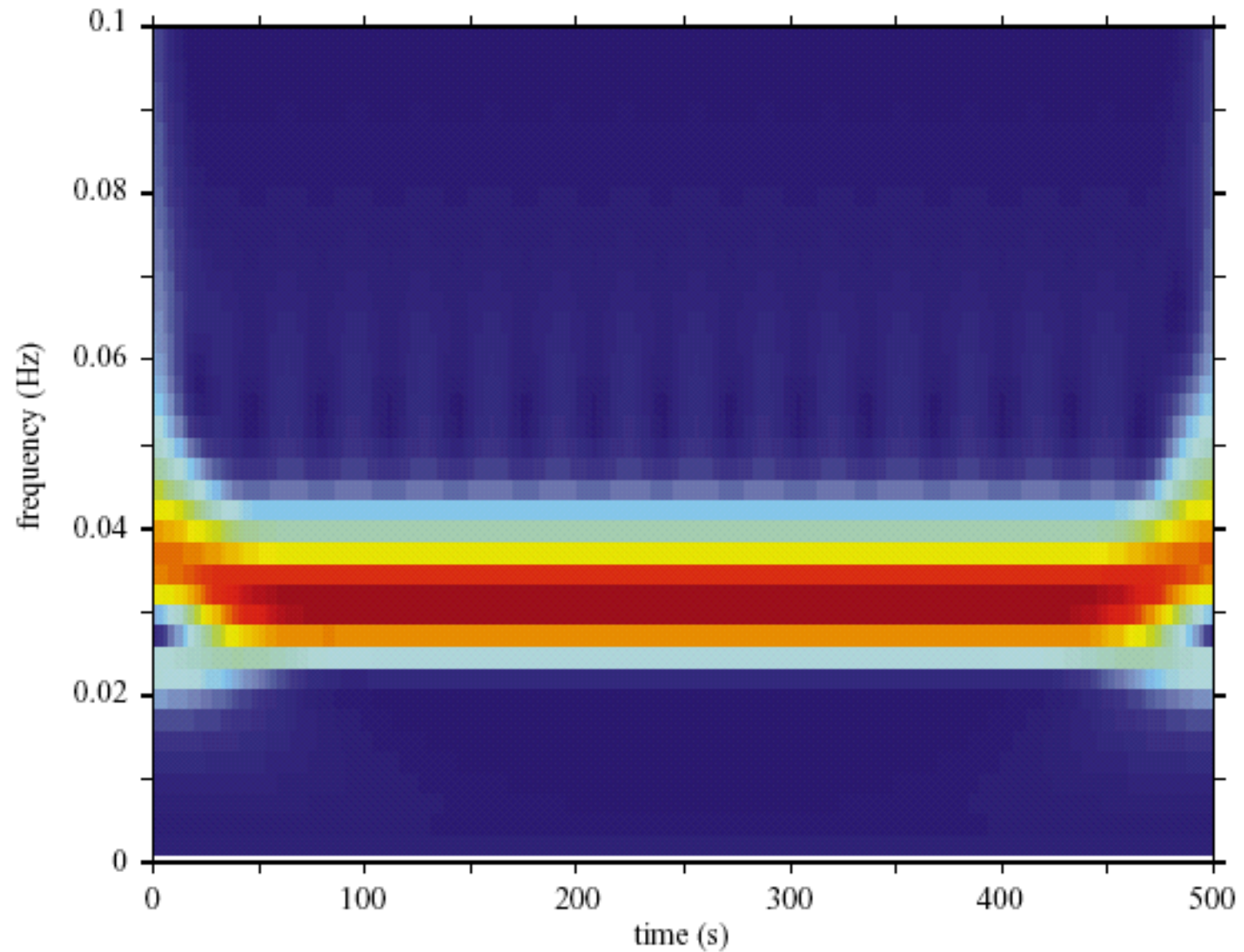


Figure 16. Morlet wavelet spectrum for the Stokes wave showing two bands of energy for the fundamental and the harmonics around 0.03 and 0.06 Hz as expected from the traditional view. The end effect of the wavelet analysis is clearly visible.

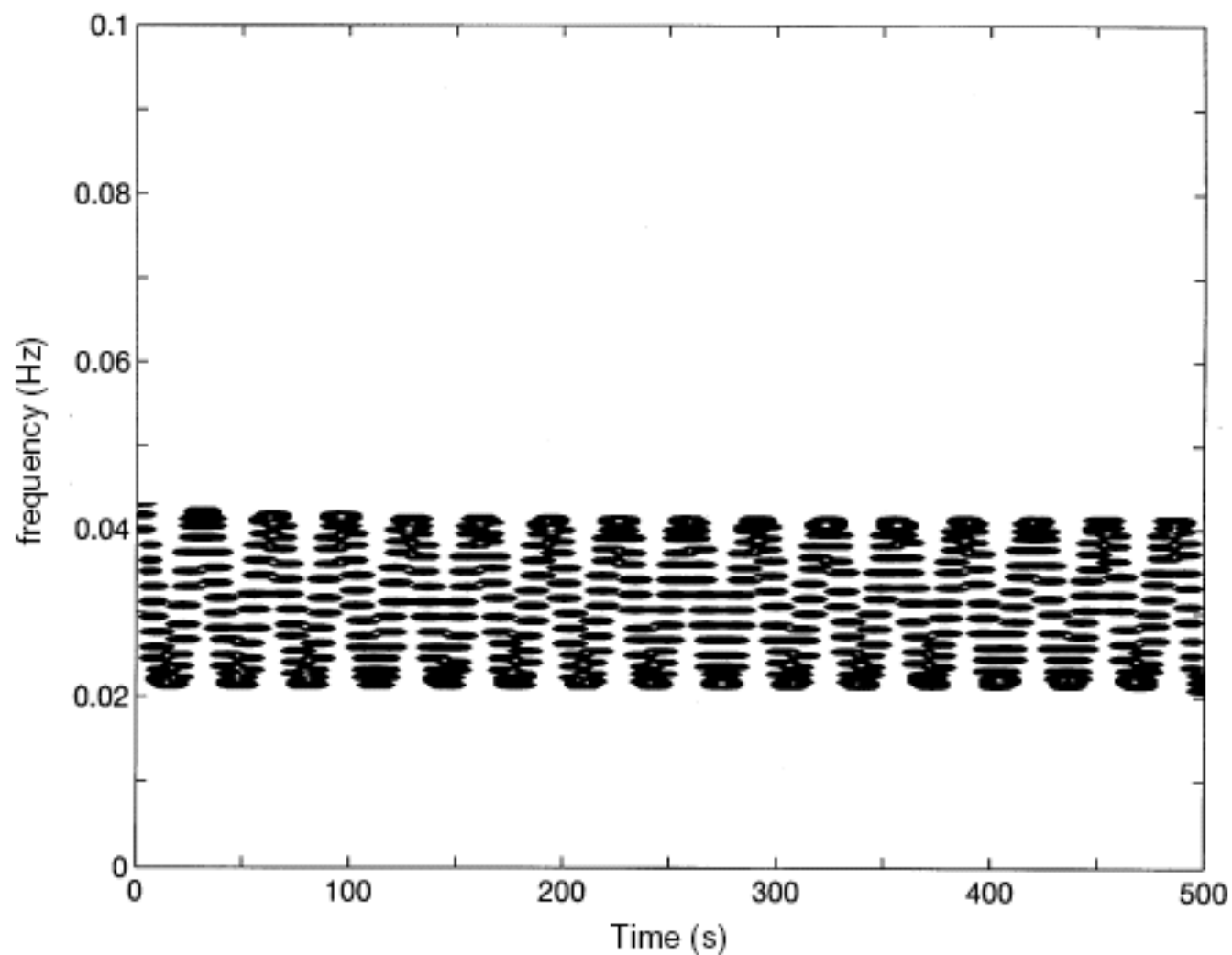


Figure 17. The Hilbert spectrum from the IMF with forced symmetry of the profile in which the intrawave frequency modulation interpretation of the Stokes wave is clear. In the Hilbert spectrum the frequency variation is bounded in a narrow range around 0.03 Hz with no harmonics.

Intrawave frequency modulation

Interpretation

- Intrawave frequency modulation (from 0.02 to 0.04 Hz) vs. harmonic distortion: is it a more physical interpretation?
 - Harmonic components often arise because a linear system is used to approximate the nonlinear one (e.g. perturbation expansion into an infinite series of Fourier modes)
 - Intrawave frequency modulation produces waves with similar distortion as the Stokes wave

Example of amplitude modulation

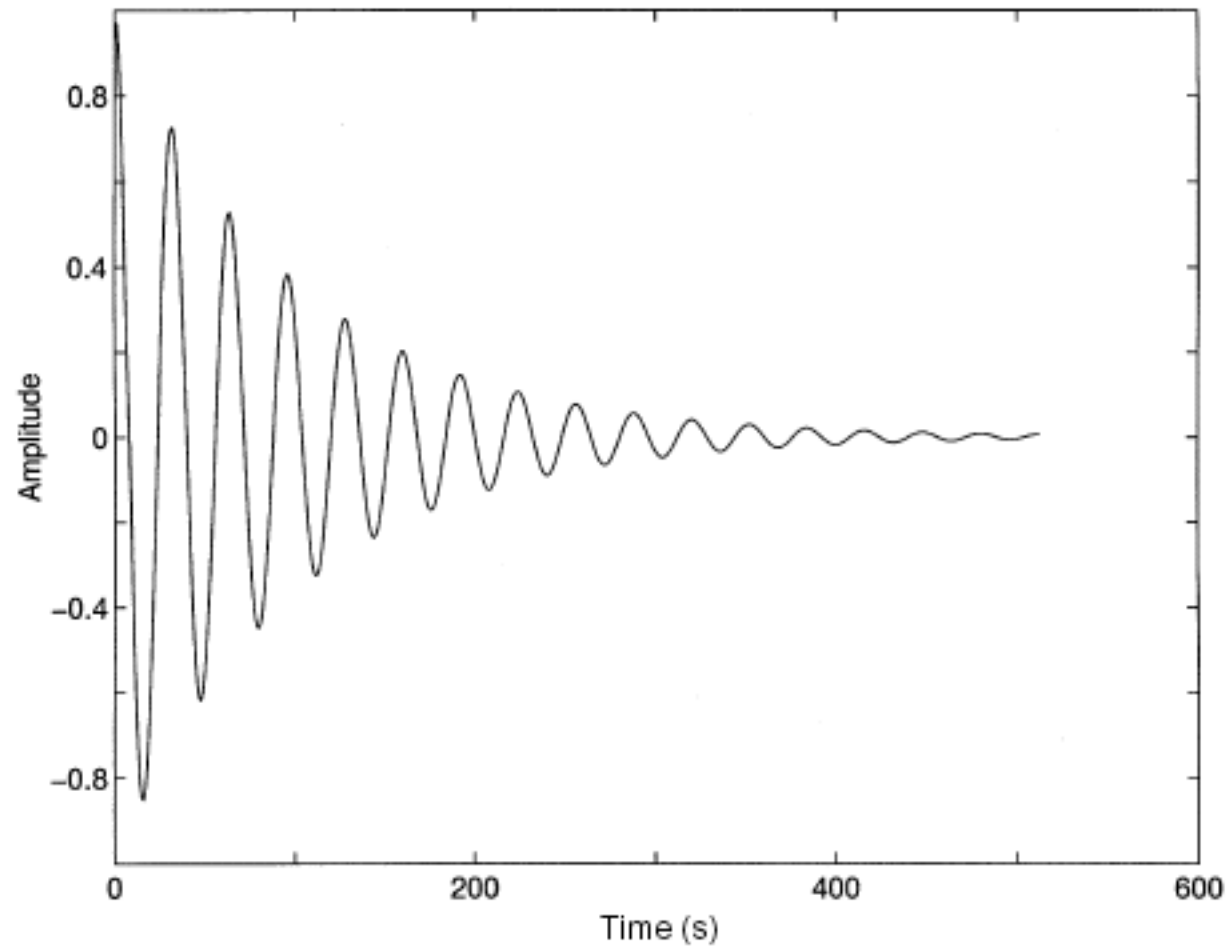


Figure 20. Data for an amplitude modulated wave: a single carrier with exponentially decaying envelope.

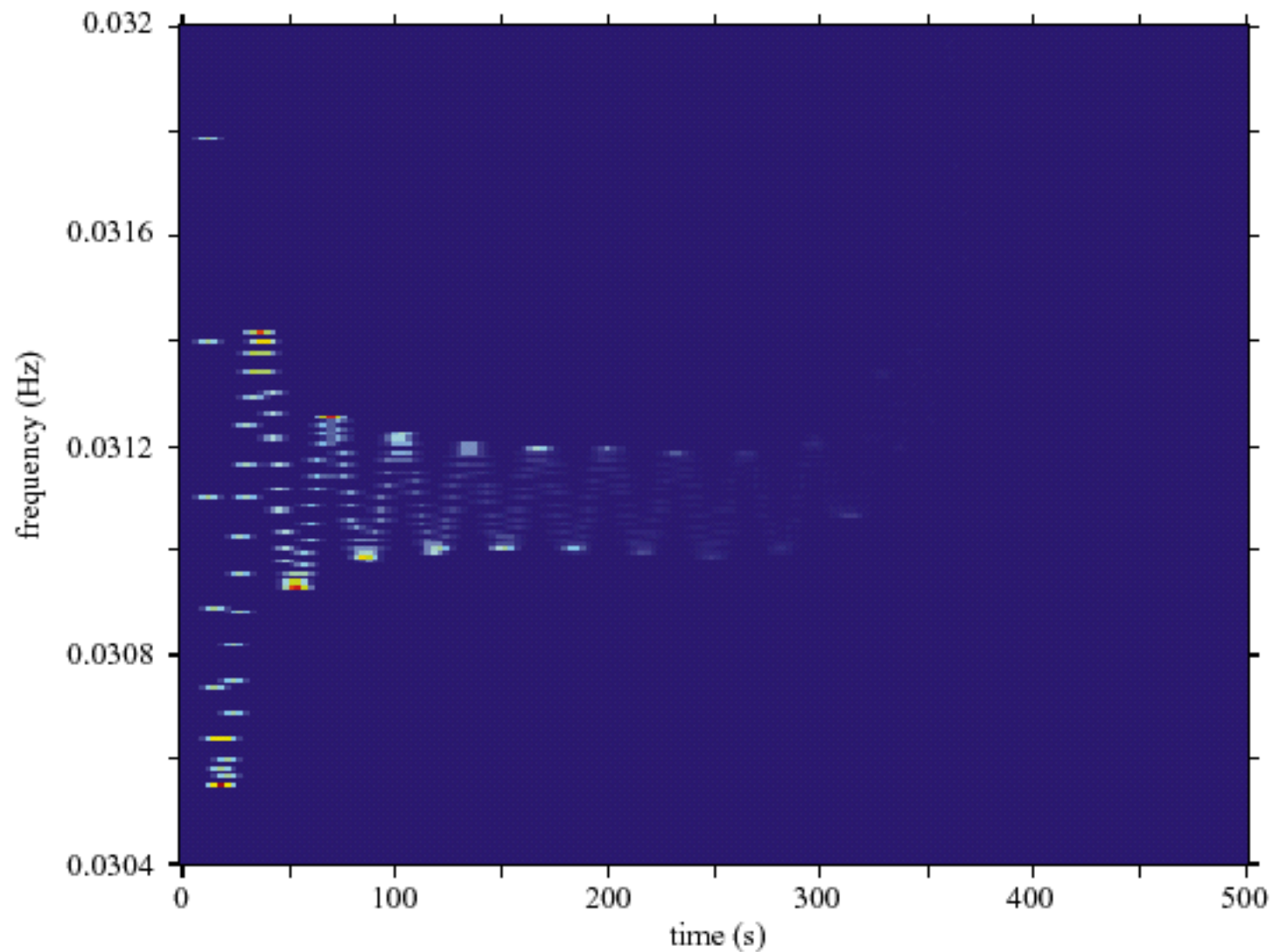


Figure 21. The Hilbert spectrum of the data shown in figure 20. This example shows that the amplitude modulation can also generate intrawave frequency modulation, but the range of the modulation is only $\pm 1.5\%$.

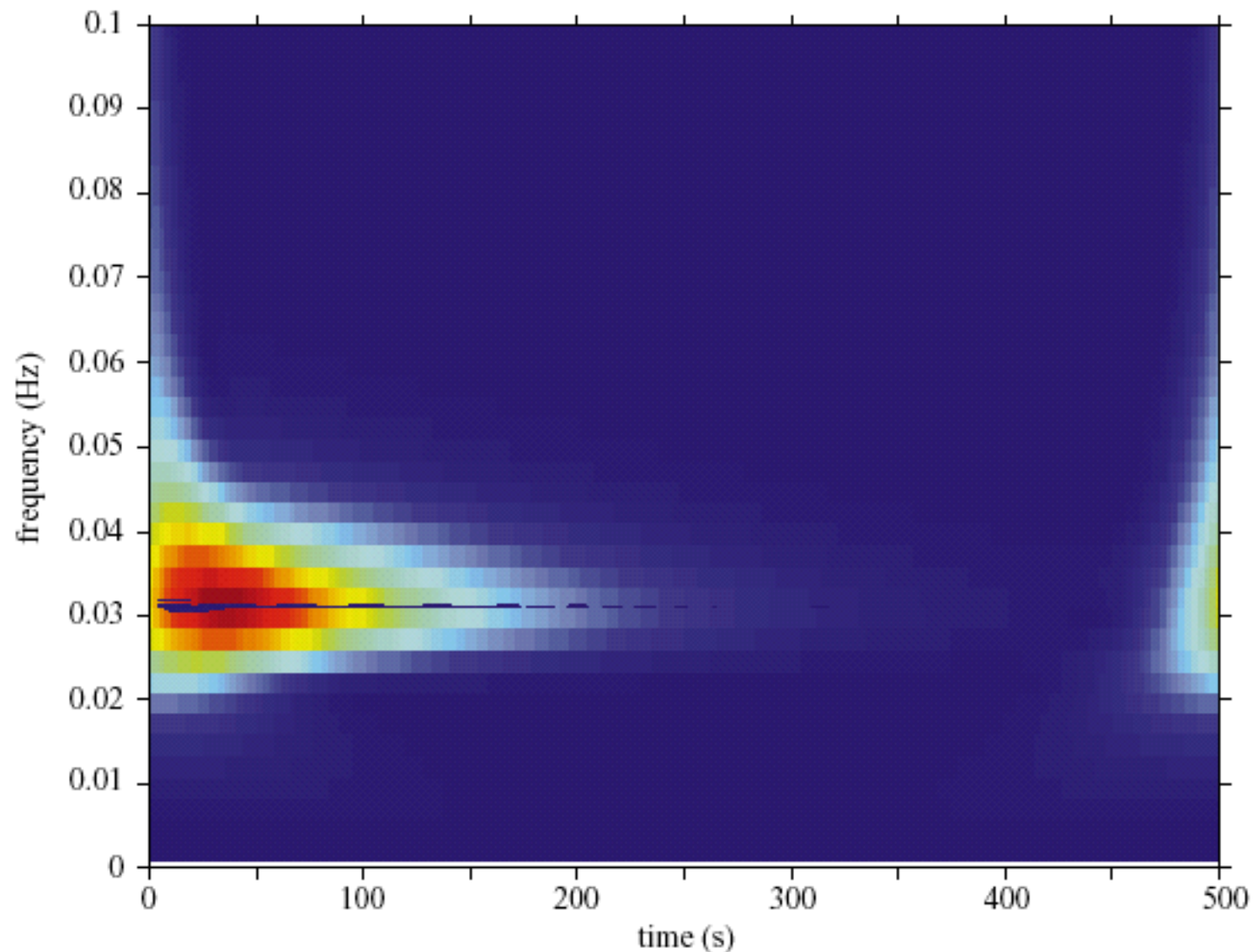


Figure 22. The wavelet spectrum from the same data shown in figure 20. In this representation, the end effect is very prominent; the frequency is widely spread. As a comparison of wavelet and Hilbert spectral analysis, the Hilbert spectrum is also plotted in this figure in contour lines, which shows up as a thin line around 0.03 Hz. This example illustrates two points: the Hilbert spectrum has much better frequency definition, and the amplitude-variation-induced frequency modulation is small.

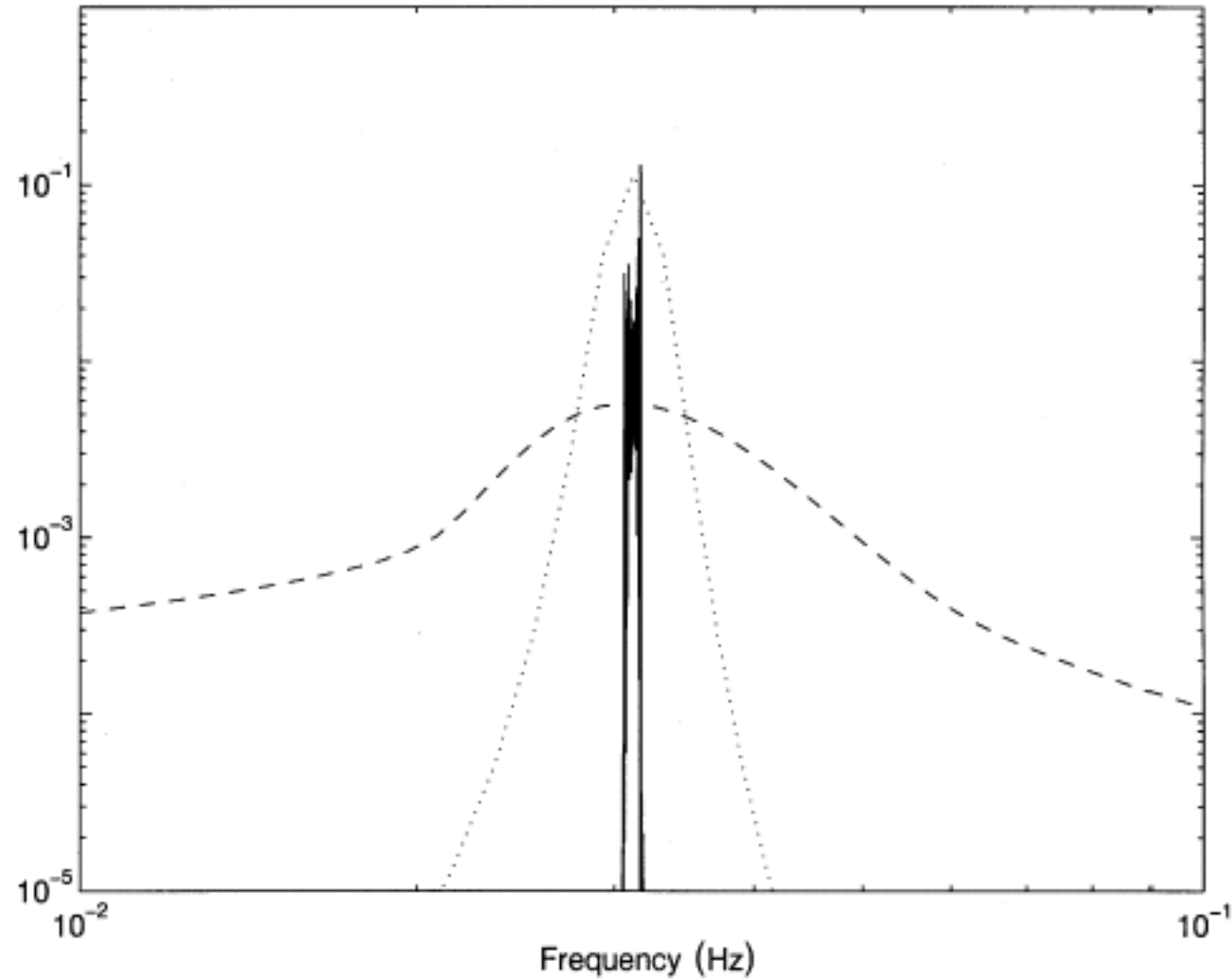
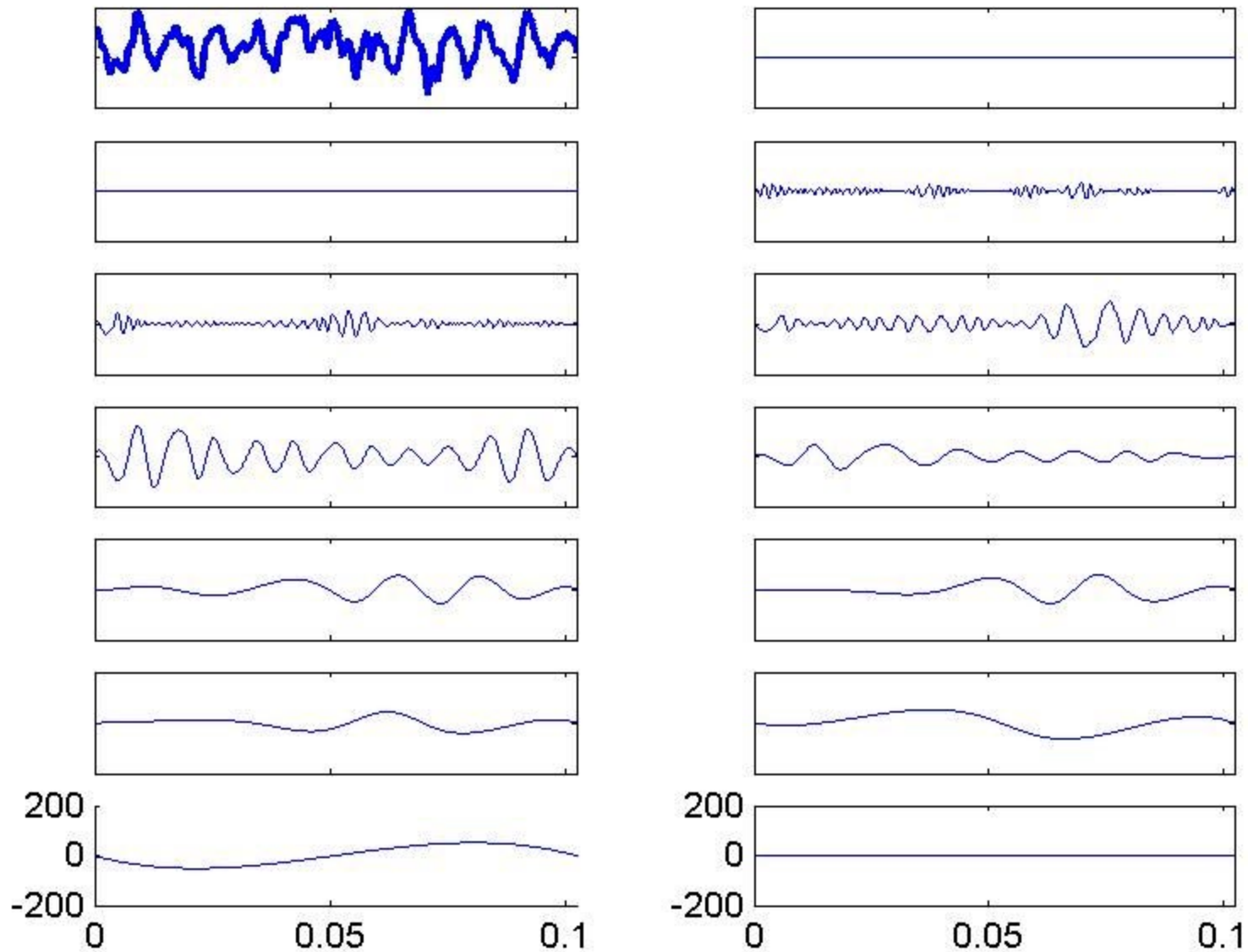


Figure 23. The comparisons among the Fourier (dotted line), marginal Hilbert (solid line) and wavelet spectra (dashed line): it is obvious that the end effects and the leakage of the wavelet analysis renders the marginal spectrum almost useless. Even the Fourier spectrum is sharper than the marginal wavelet spectrum, but it totally failed to show the time variation of the signal. The marginal Hilbert is still sharp to define the carrier frequency.

Neuroscience examples: LFP

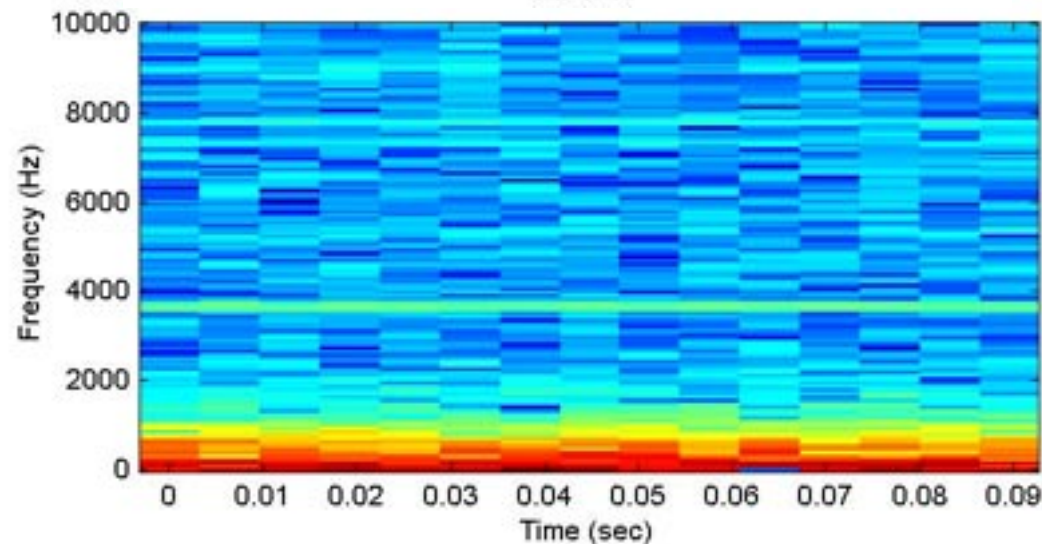
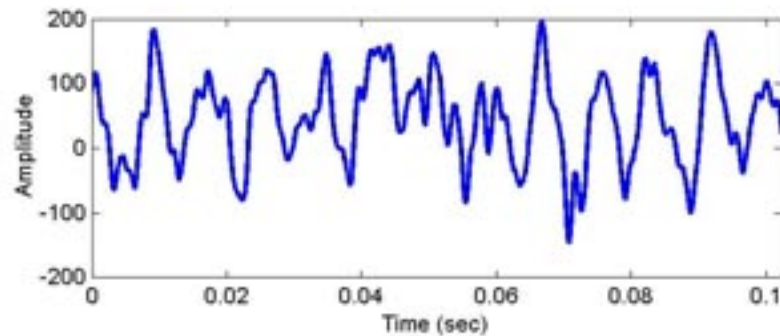
Original LFP



Fourier methods

If piece-wise stationarity can be assumed

- Construct spectrogram by sliding a window across the time-series and performing Fourier analysis

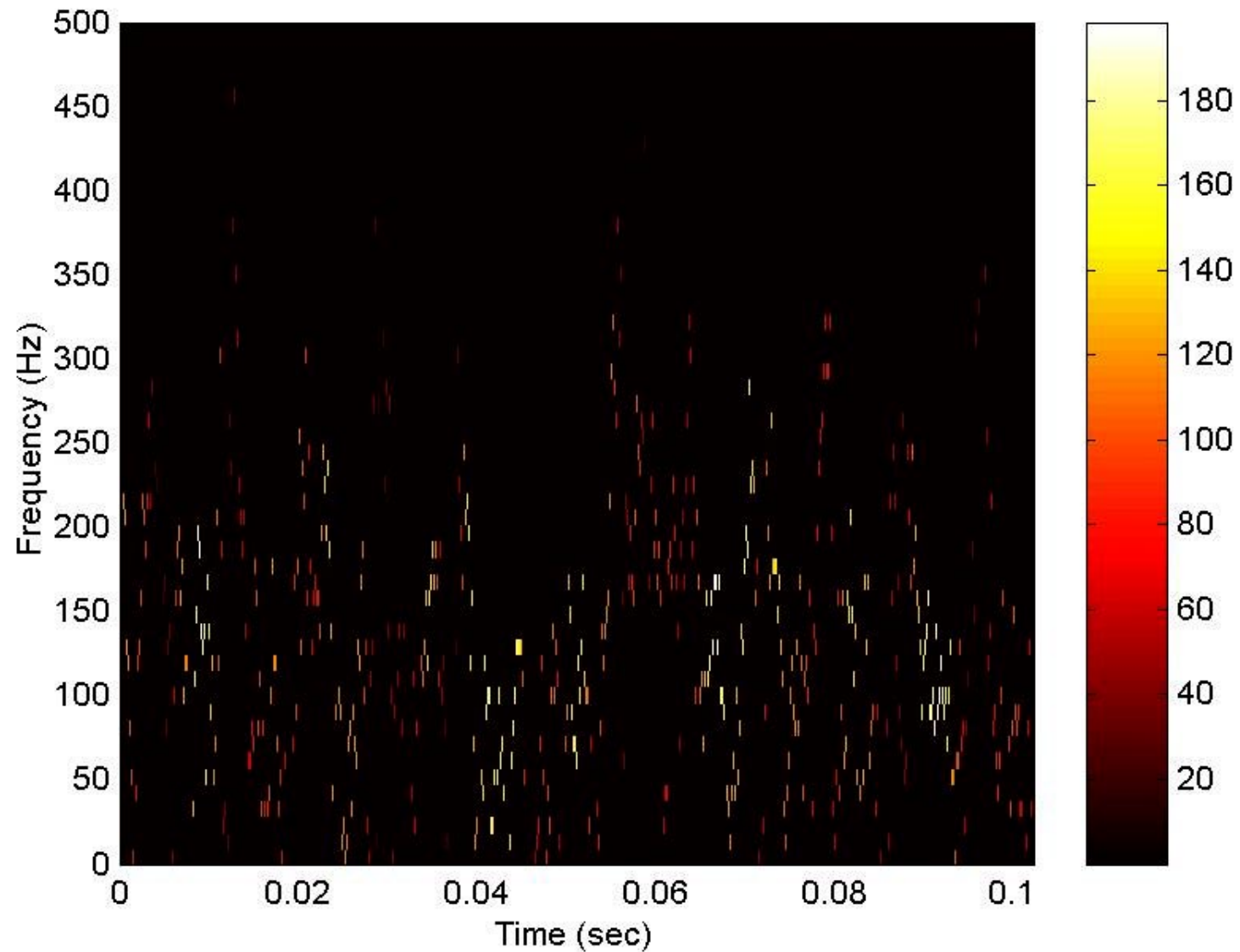


- Cor
ter

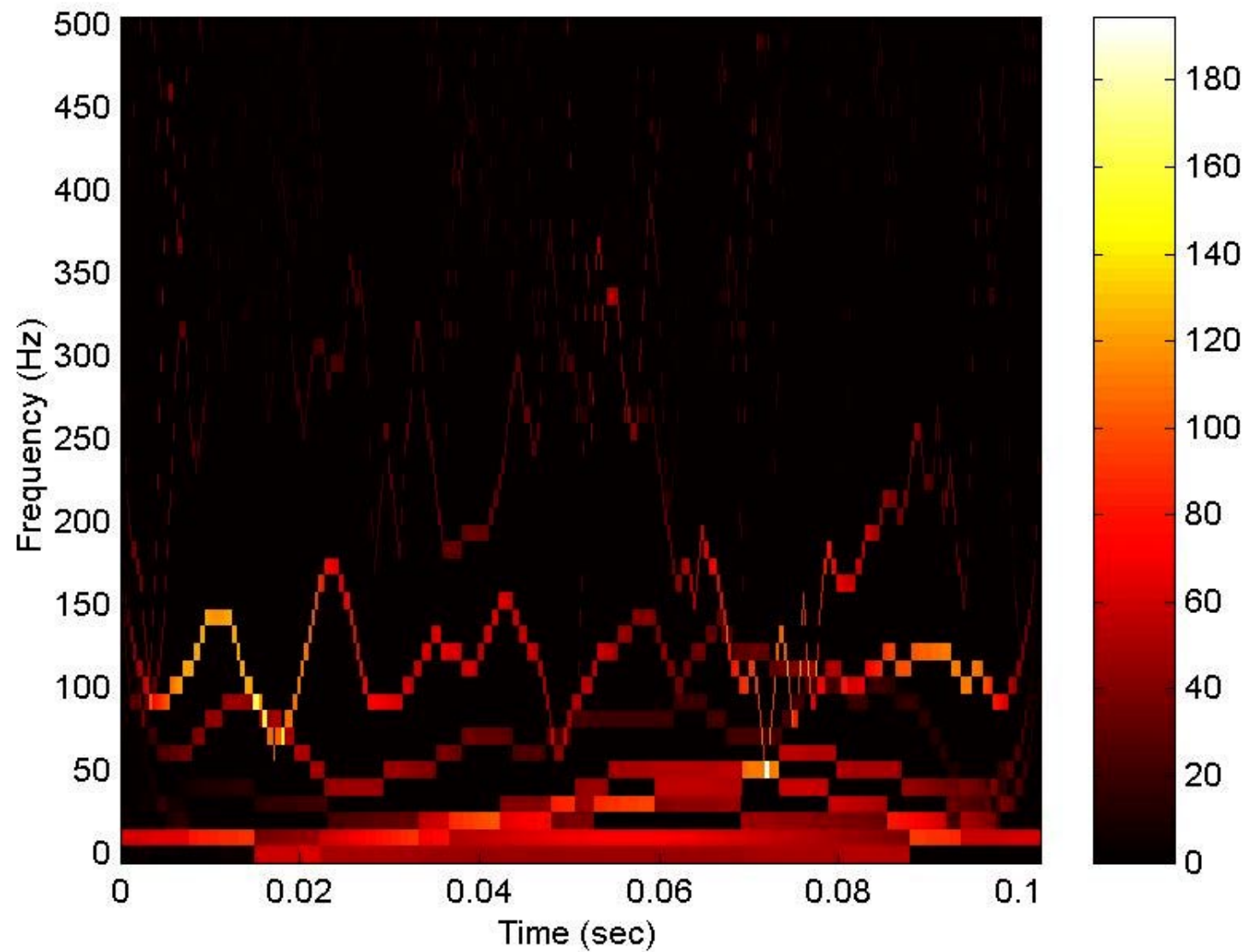
pture the
nonlocal

Hilbert spectrum of the LFP

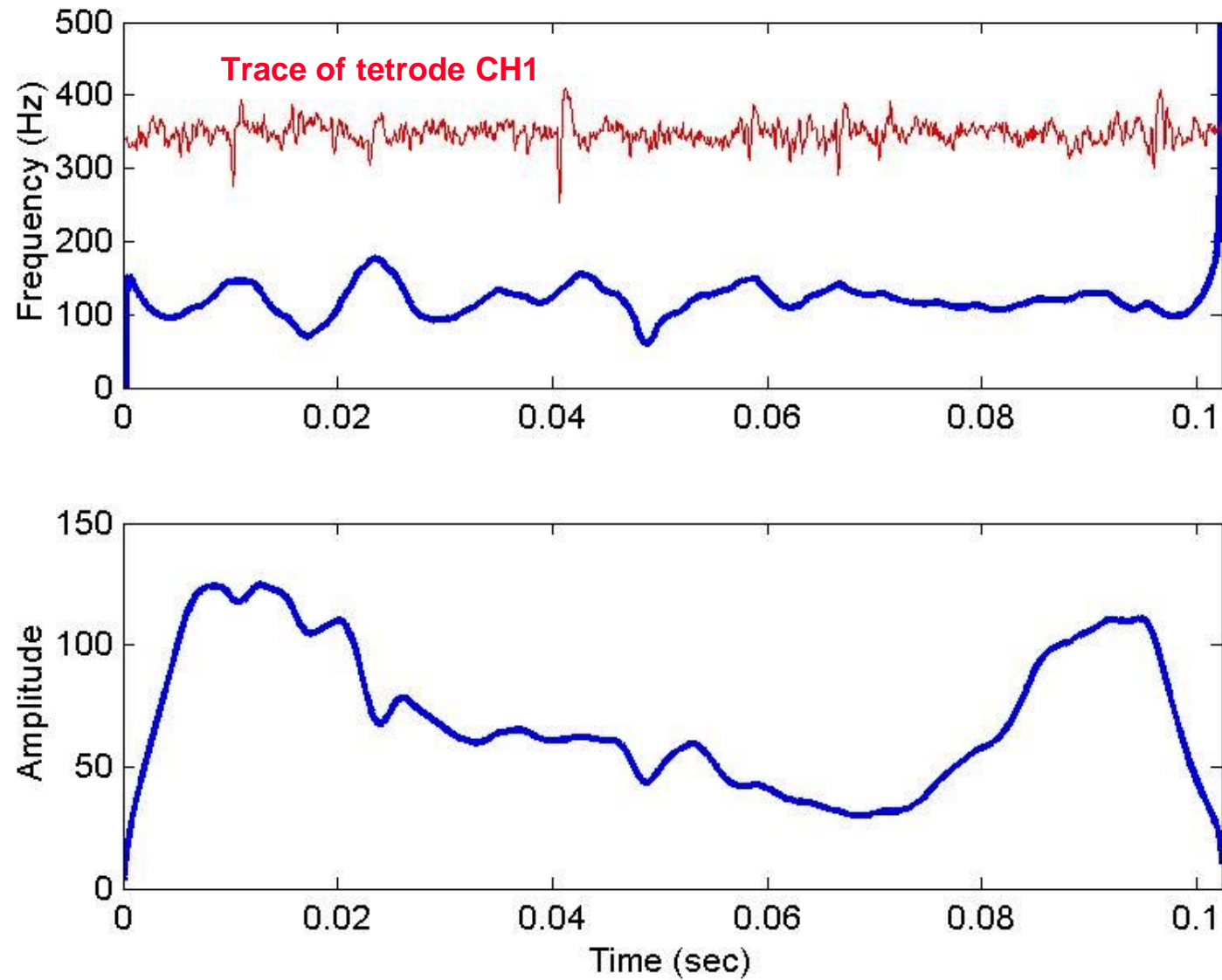
Not so useful. . . .



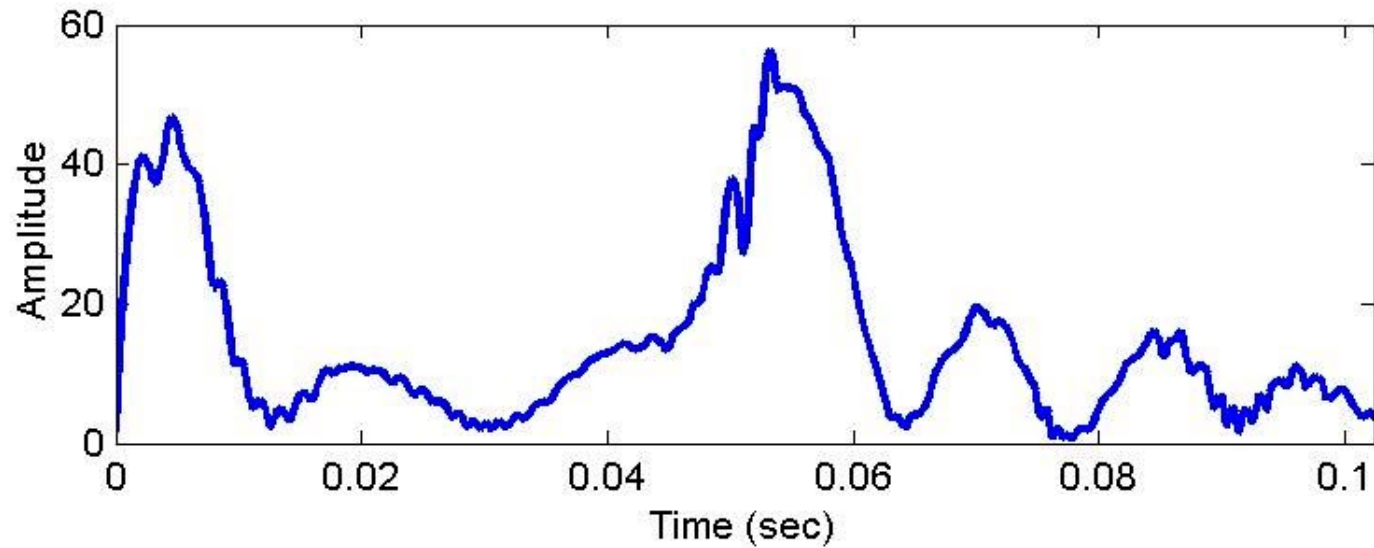
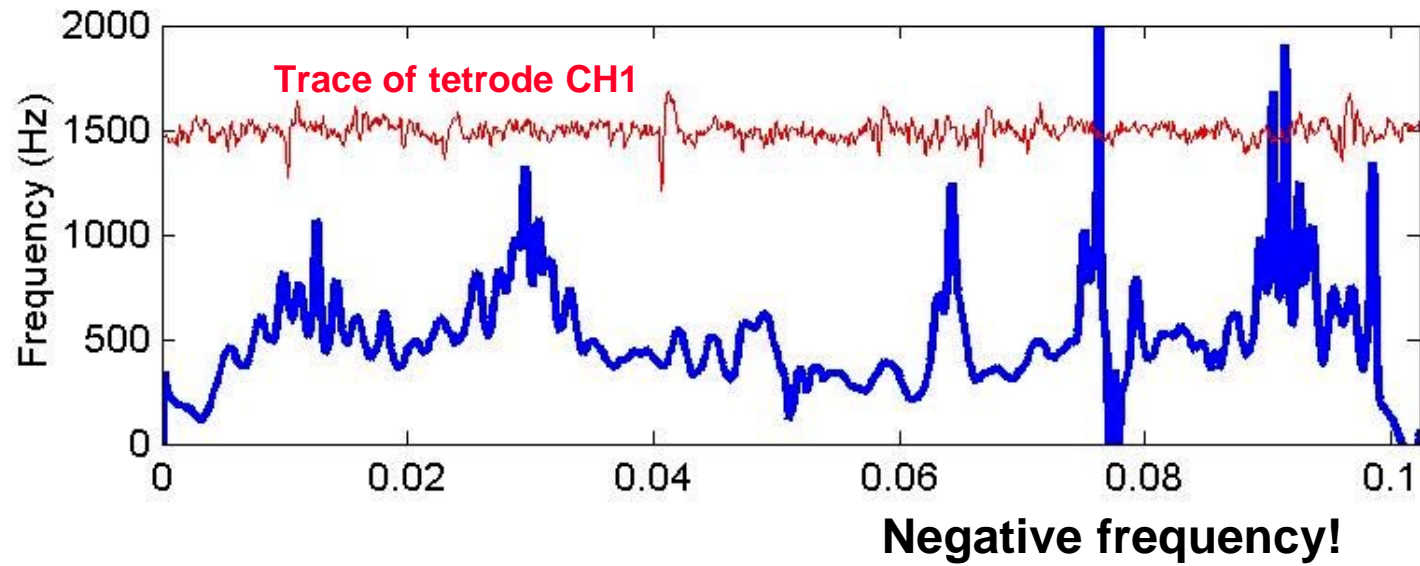
Hilbert spectrum of LFP IMF modes



Focusing on IMF mode 6

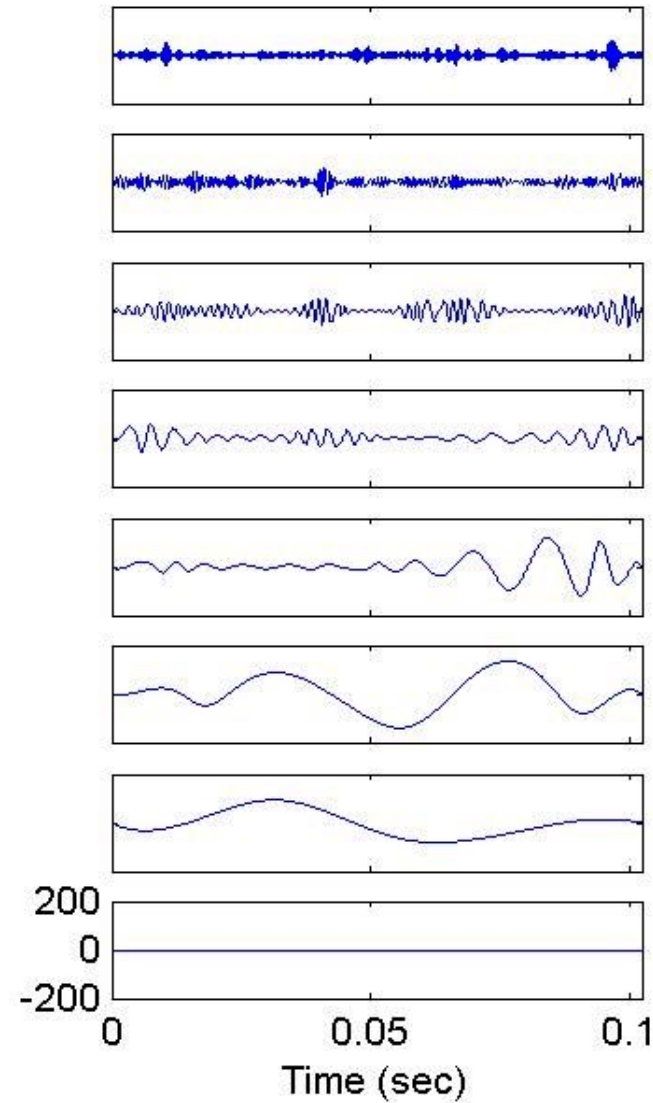
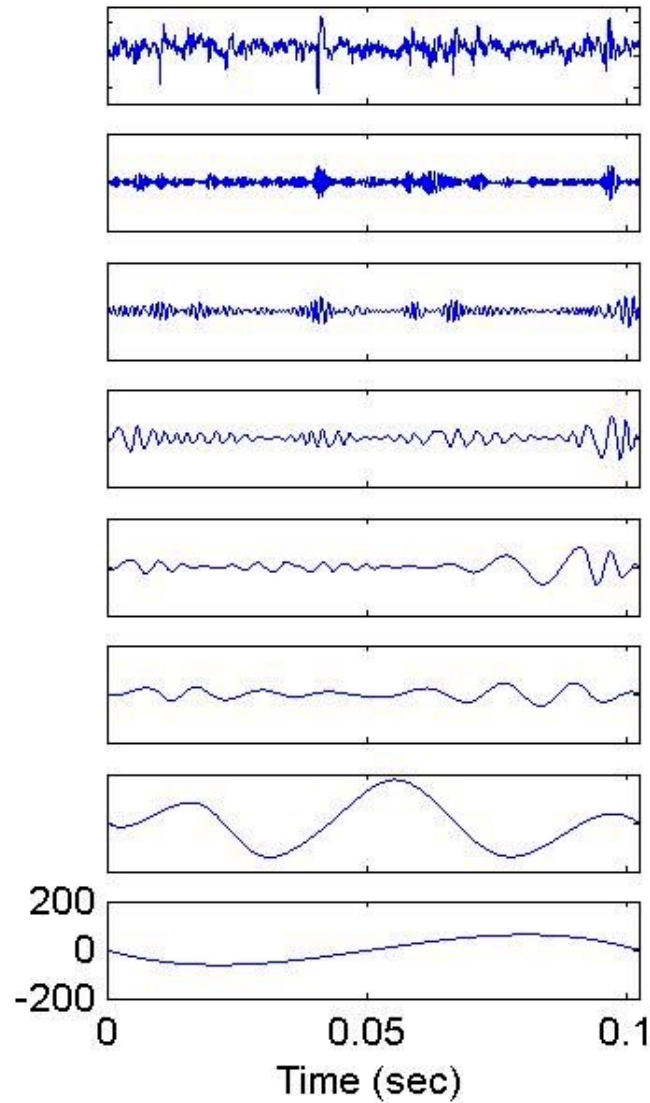


Focusing on IMF mode 4

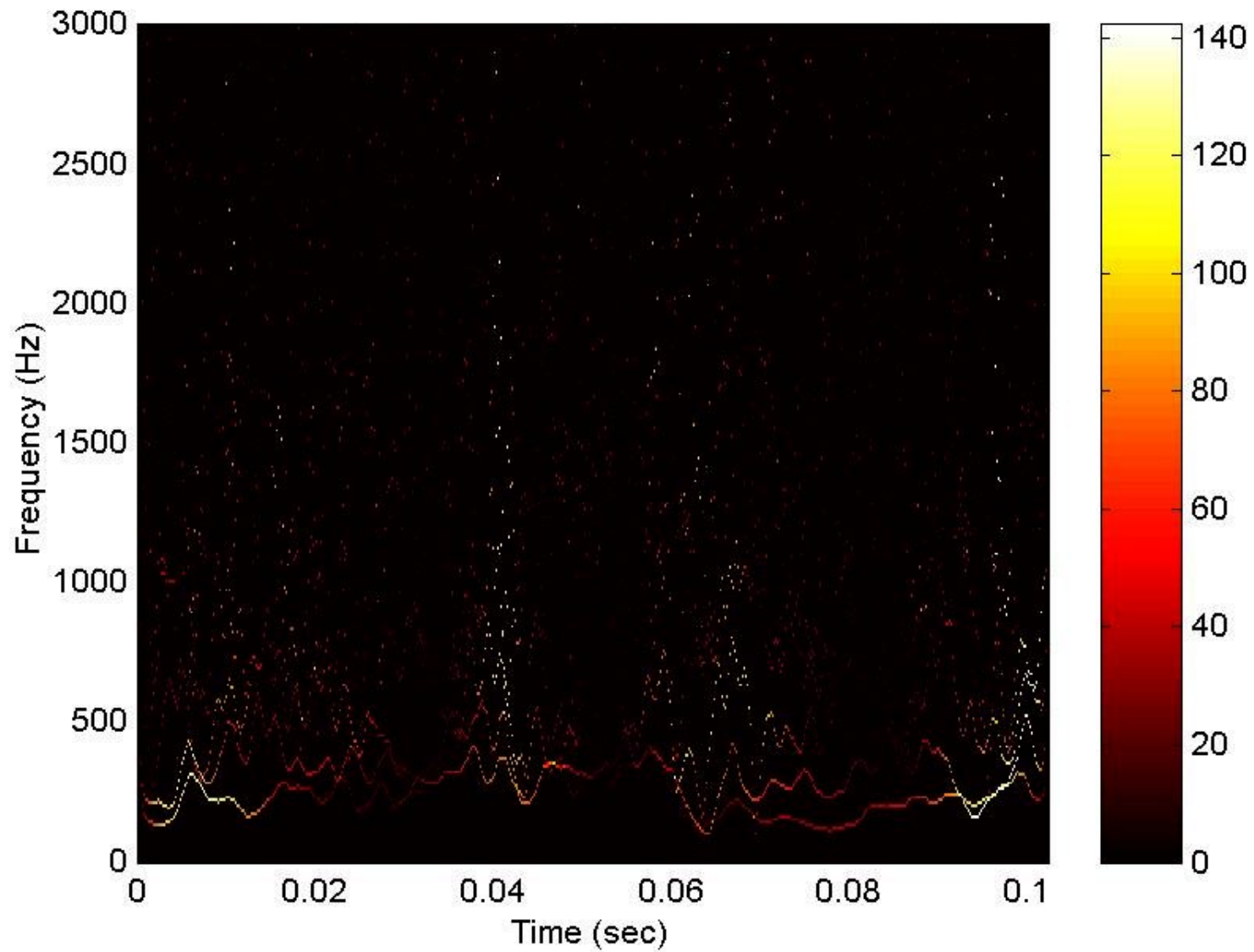


Tetrode CH1 modes

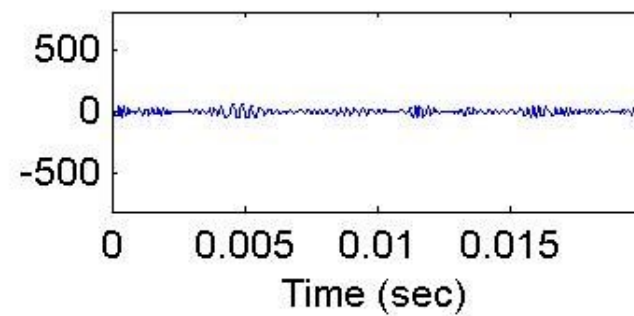
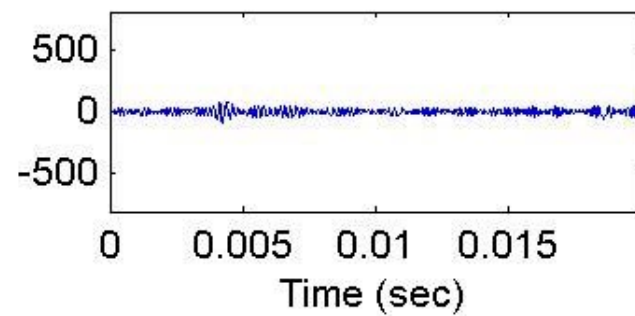
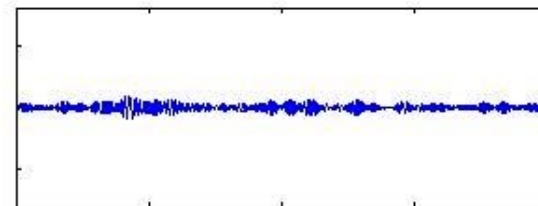
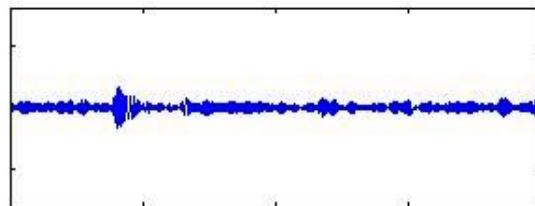
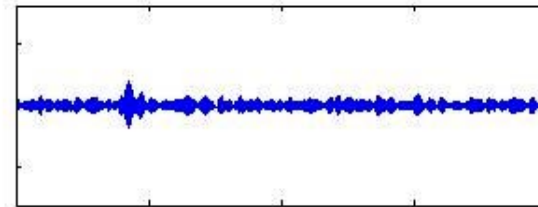
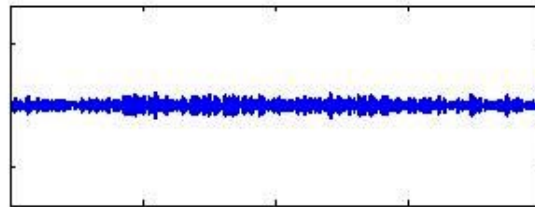
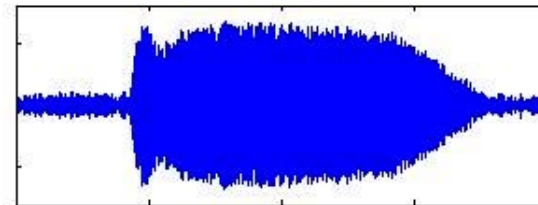
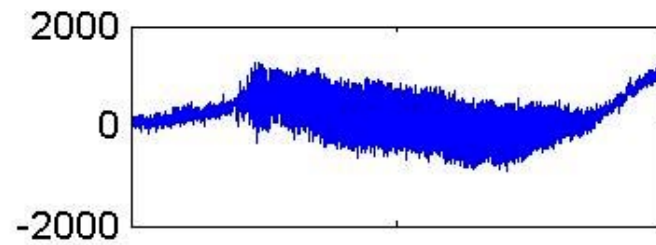
Original CH1 data



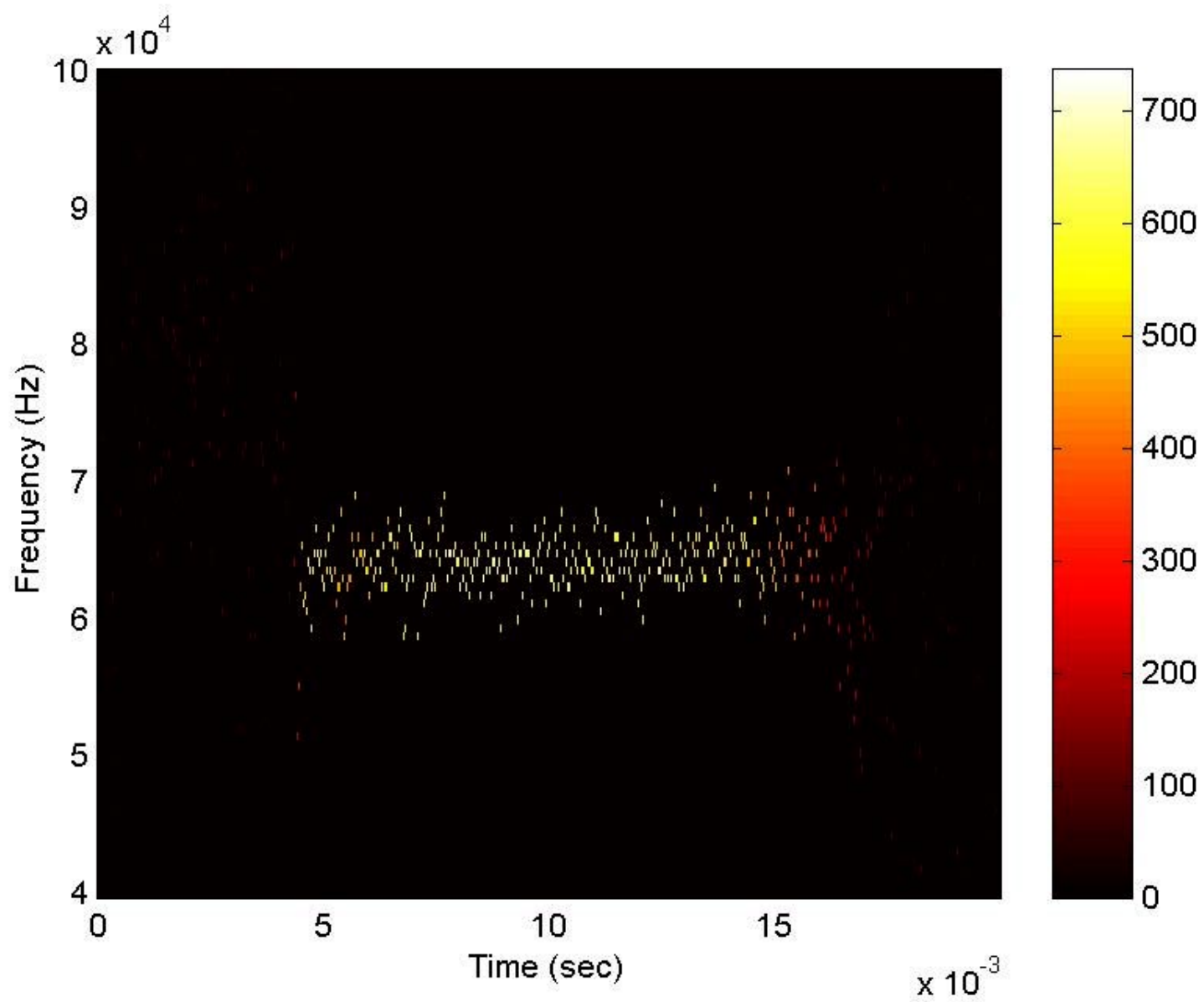
Hilbert spectrum of IMF modes 1-7



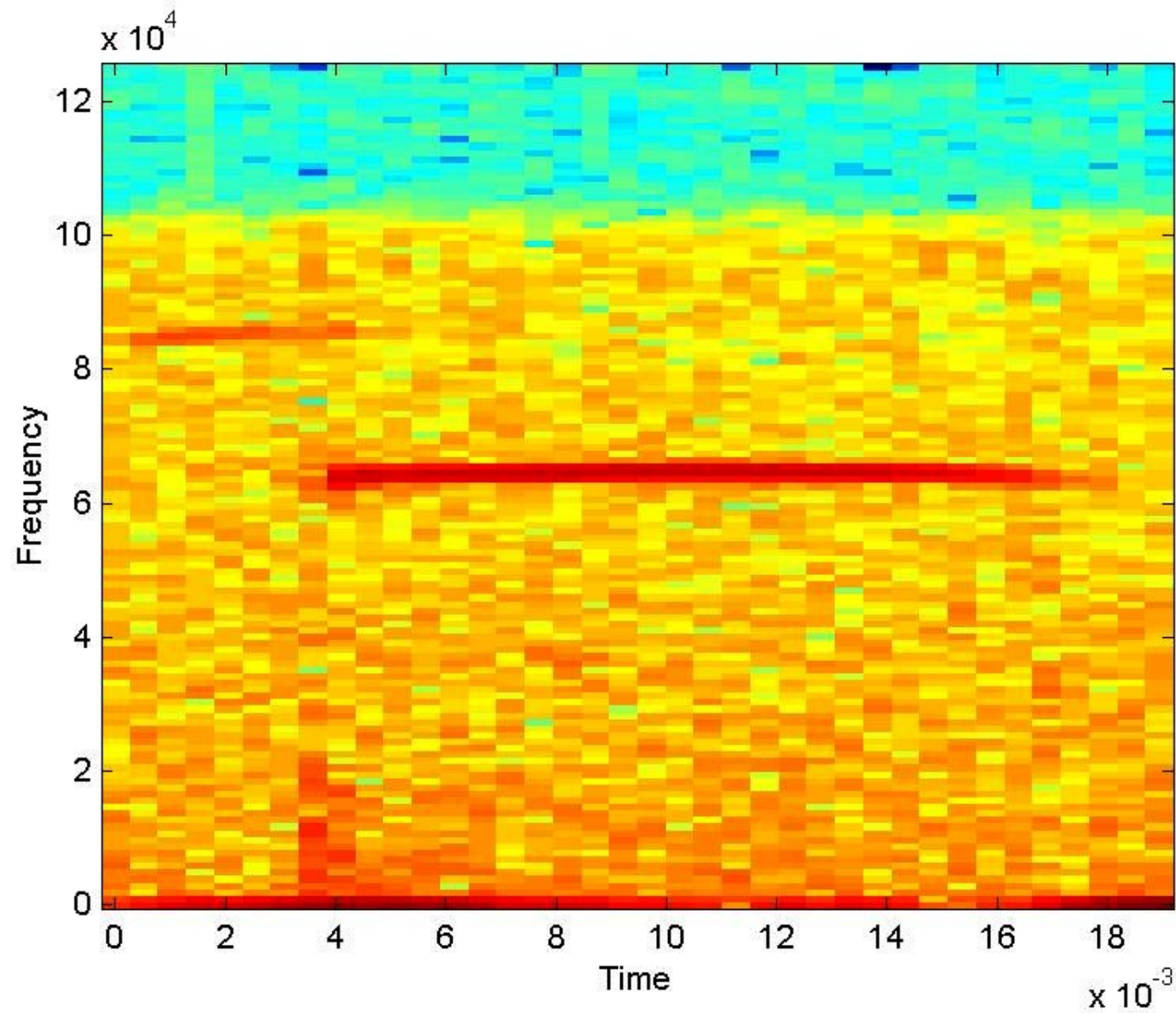
Mouse call



Hilbert spectrum of IMF mode 1



Spectrogram of mouse call



Conclusions

- **Novel technique for analyzing non-stationary and nonlinear time series**
- **Best suited for cases when signal time scales (and energy?) are distinct from the noise time scales**
- **Best suited for situations where oscillations may provide a physical description**
- **Technical improvements are necessary (but have been developed over the last 5 years)**